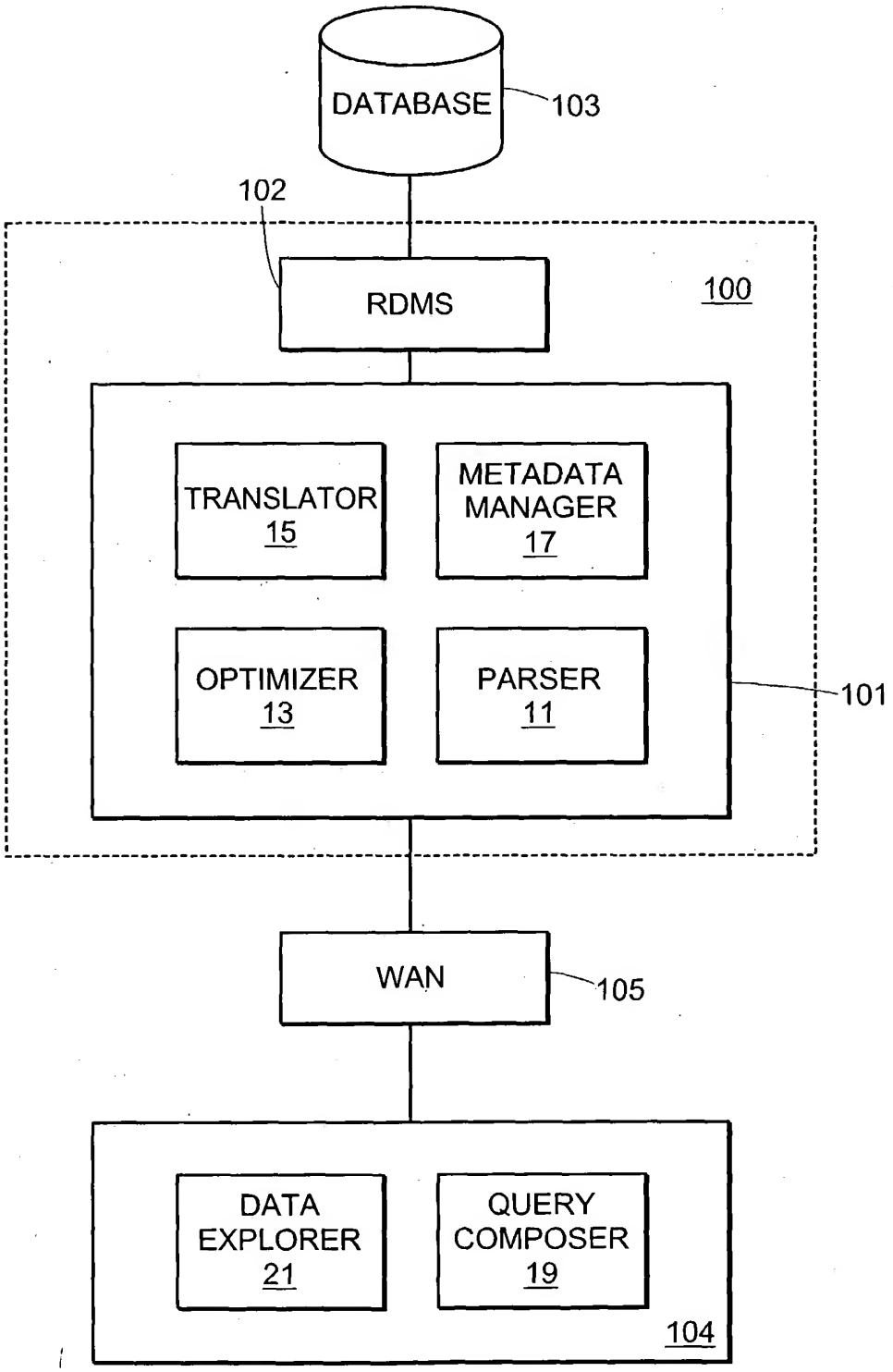
DATABASE
103
102
RDMS
100
TRANSLATOR
15
METADATA
MANAGER
17
OPTIMIZER
13
PARSER
11
101
WAN
105
DATA
EXPLORER
21
QUERY
COMPOSER
19
104

FIG. 1

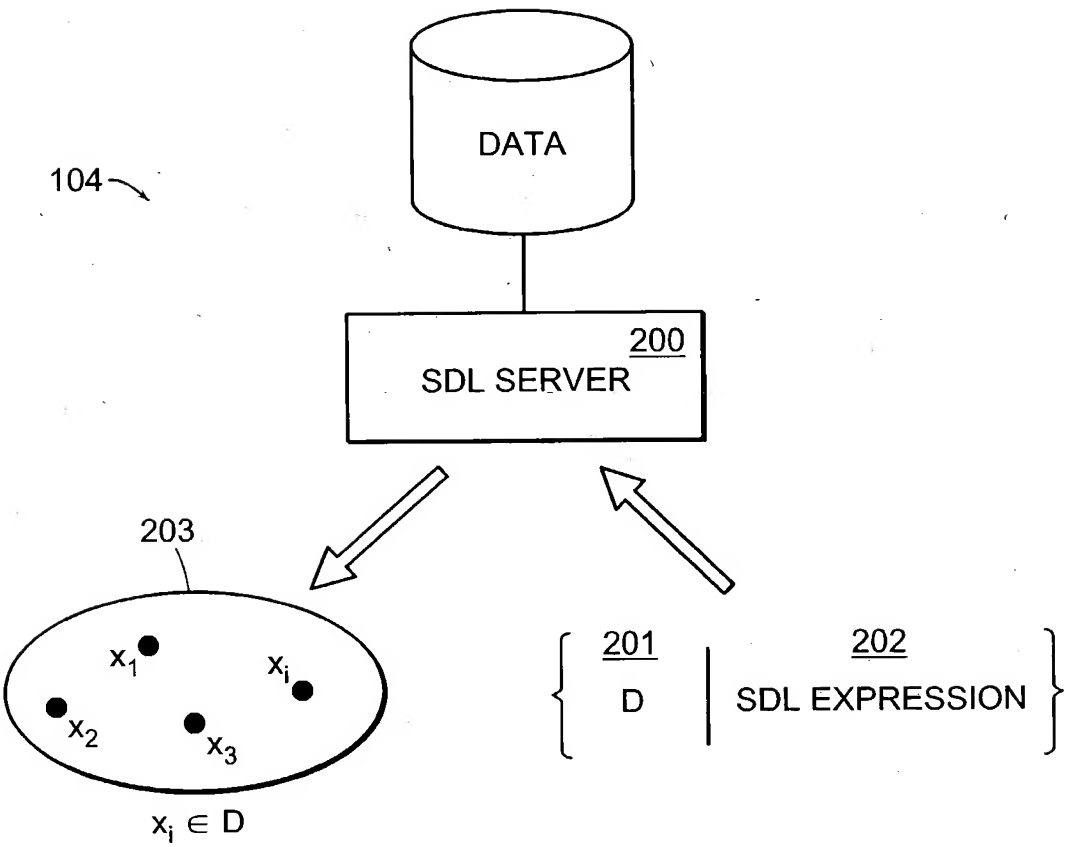
104
DATA
200
SDL SERVER
203
x_1
x_i
x_2
x_3
$x_i \in D$
201
D
202
SDL EXPRESSION

FIG. 2

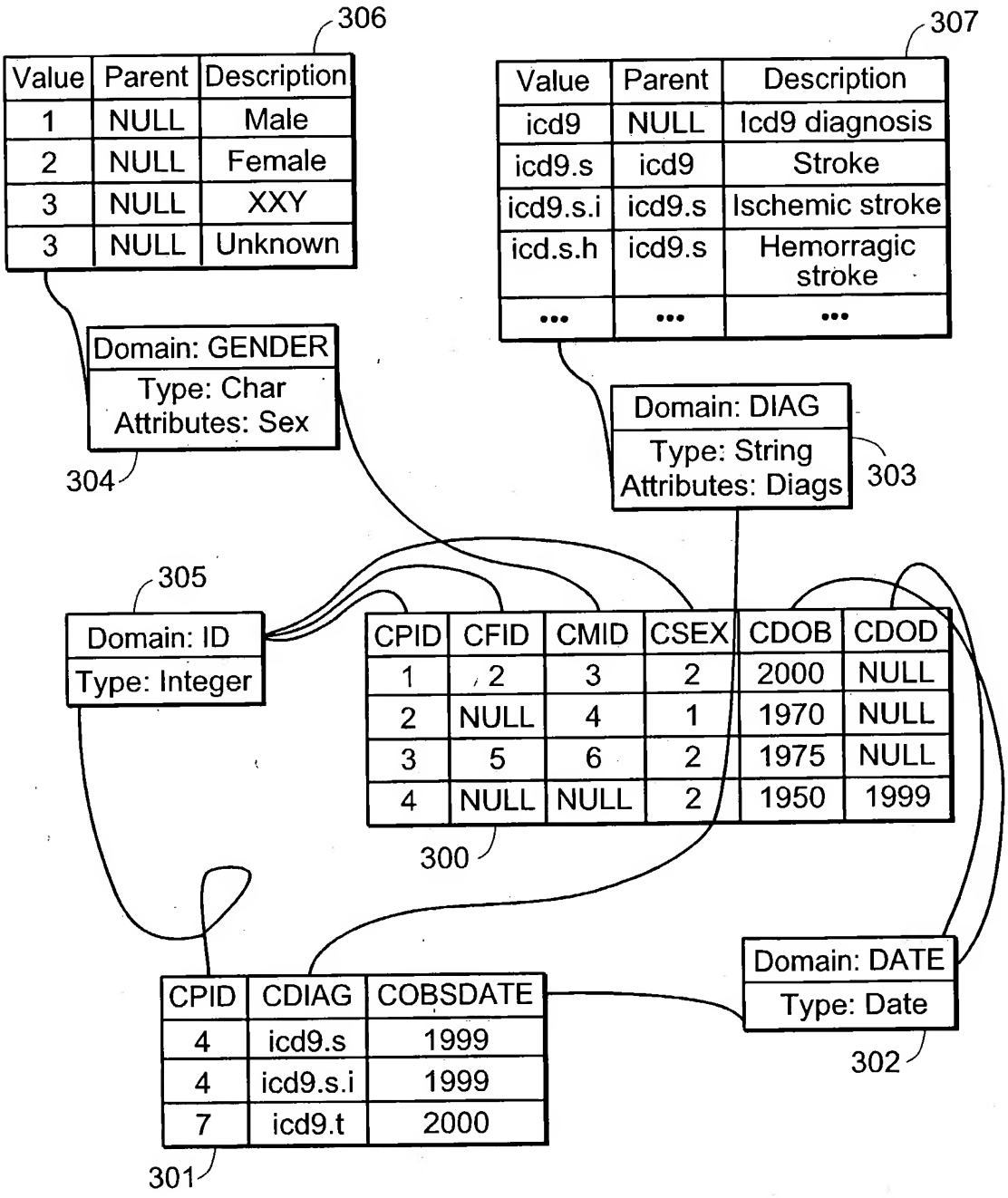
306
Value | Parent | Description
1 | NULL | Male
2 | NULL | Female
3 | NULL | XXY
3 | NULL | Unknown
307
Value | Parent | Description
icd9 | NULL | Icd9 diagnosis
icd9.s | icd9 | Stroke
icd9.s.i | icd9.s | Ischemic stroke
icd.s.h | icd9.s | Hemorragic stroke
... | ... | ...
Domain: GENDER
Type: Char
Attributes: Sex
304
Domain: DIAG
Type: String
Attributes: Diags
303
305
Domain: ID
Type: Integer
CPID | CFID | CMID | CSEX | CDOB | CDOD
1 | 2 | 3 | 2 | 2000 | NULL
2 | NULL | 4 | 1 | 1970 | NULL
3 | 5 | 6 | 2 | 1975 | NULL
4 | NULL | NULL | 2 | 1950 | 1999
300
CPID | CDIAG | COBSDATE
4 | icd9.s | 1999
4 | icd9.s.i | 1999
7 | icd9.t | 2000
301
Domain: DATE
Type: Date
302

FIG. 3

400

	Domain	SQLtype	Hierarchy	Min	Max
303	DIAG	String	TDiags	NULL	NULL
304	GENDER	Char	TSex	NULL	NULL
302	DATE	Date	NULL	1900	2100
305	ID	Integer	NULL	0	NULL

401

Dimension	Domain
PID	ID
FID	ID
MID	ID
SEX	GENDER
DOB	DATE
DOD	DATE
OBSDATE	DATE
DIAG	DIAG

403

Dimension	Relation	Column_name	Multiplicity
PID	relation1	CPID	Unique
PID	relation2	CPID	Multiple
FID	relation1	CFID	Multiple
MID	relation1	CMID	Multiple
SEX	relation1	CSEX	Multiple
DOB	relation1	CDOB	Multiple
DOD	relation1	CDOD	Multiple
OBSDATE	relation2	COBSDATE	Multiple
DIAG	relation2	CDIAG	Multiple

402

Relation	SQLrelation	Inclusion_criteria
relation1	TableA	{PID, FID, MID, SEX, DOB, DOD}
relation2	TableB	Sex

FIG. 4

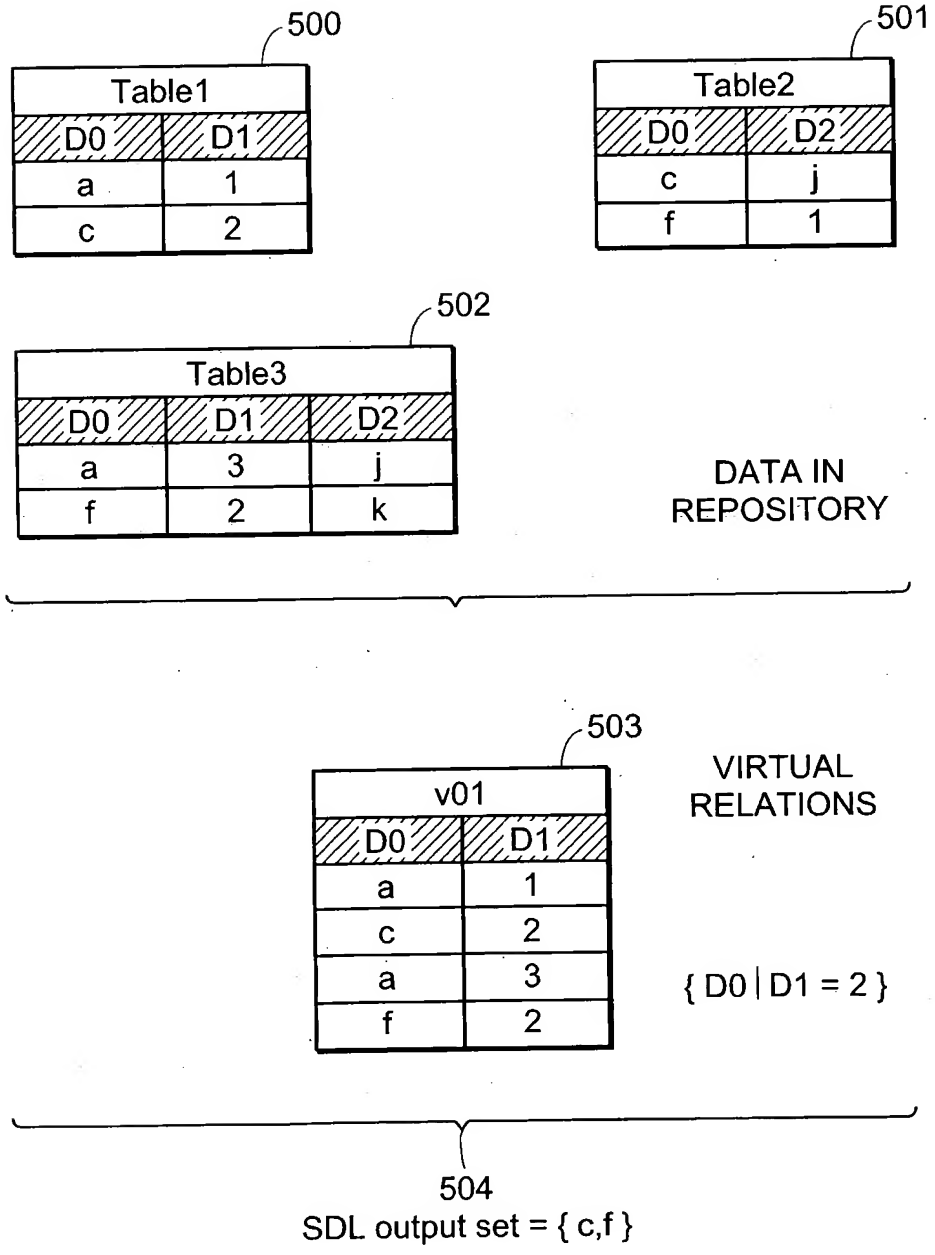
500
Table1
D0
D1
a
1
c
2
501
Table2
D0
D2
c
j
f
1
502
Table3
D0
D1
D2
a
3
j
f
2
k
DATA IN REPOSITORY
503
v01
D0
D1
a
1
c
2
a
3
f
2
VIRTUAL RELATIONS
{ D0 | D1 = 2 }
504
SDL output set = { c,f }

FIG. 5

600

PID	DIAG	HOSPID	DOCID	DATE
1	icd9.stroke	100	A	1966
2	icd9.x	200	B	2000

601

HOSPID	NAME	TYPE	REGION	ZIP
100	MGH	TH	Boston	02222
200	BWH	TH	Brookline	02115

602

PID	DATE	DIAG	HOSP. NAME	HOSP. TYPE	HOSP. REGION	HOSP. ZIP
1	1966	icd9.stroke	MGH	TH	Boston	02222
2	2000	icd9.x	BWH	TH	Brookline	02115

603

Relation	SQL relation	Inclusion_criteria
relation1	diags	{PID,DIAG,HOSPID,DOCID}
relation2	hospitals	{HOSPID,NAME,TYPE,REGION, ZIP}
relation3	vdiags	{PID,DIAG,HOSPID,DOCID}, {HOSP.NAME,HOSP.TYPE, HOSP.REGION,HOSP.ZIP}

604

FIG. 6

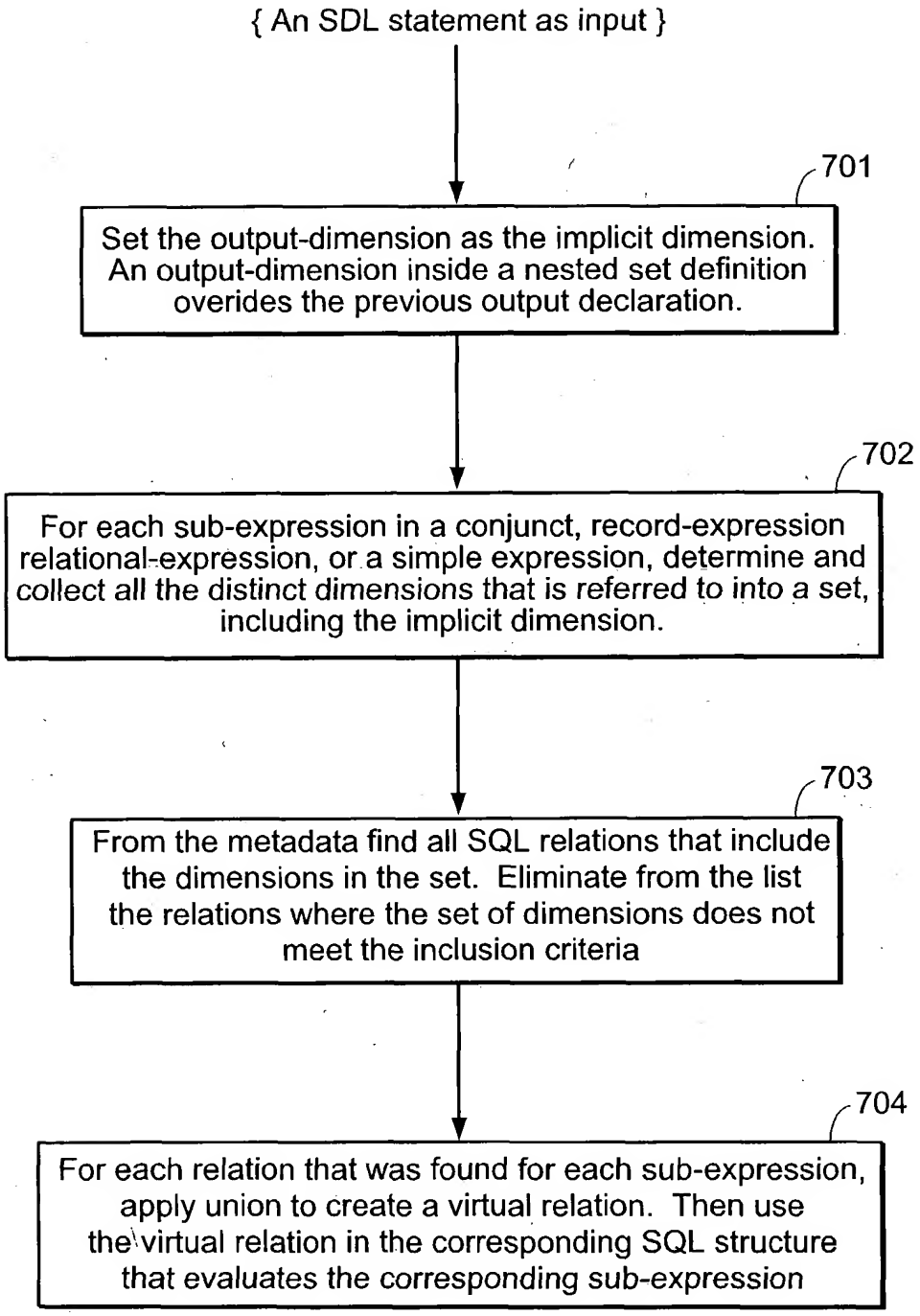
{ An SDL statement as input }
701
Set the output-dimension as the implicit dimension. An output-dimension inside a nested set definition overides the previous output declaration.
702
For each sub-expression in a conjunct, record-expression relational-expression, or a simple expression, determine and collect all the distinct dimensions that is referred to into a set, including the implicit dimension.
703
From the metadata find all SQL relations that include the dimensions in the set. Eliminate from the list the relations where the set of dimensions does not meet the inclusion criteria
704
For each relation that was found for each sub-expression, apply union to create a virtual relation. Then use the virtual relation in the corresponding SQL structure that evaluates the corresponding sub-expression

FIG. 7

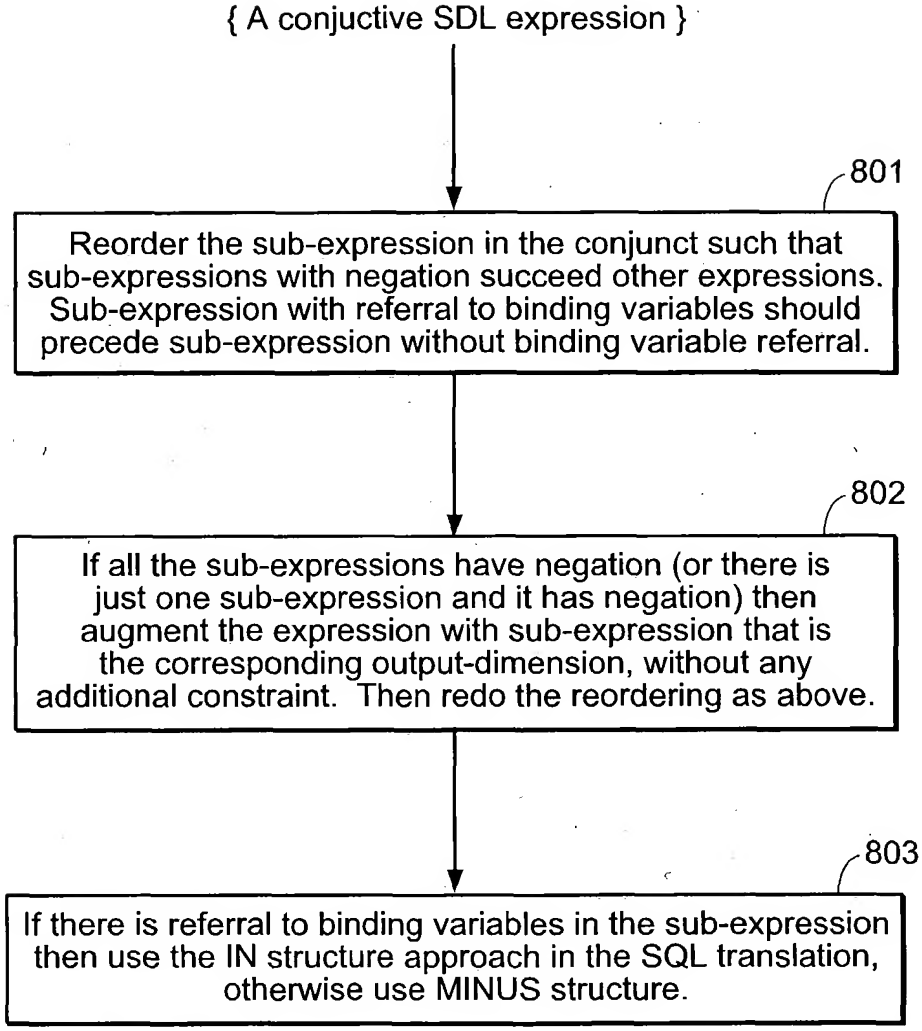
{ A conjuctive SDL expression }
801
Reorder the sub-expression in the conjunct such that sub-expressions with negation succeed other expressions. Sub-expression with referral to binding variables should precede sub-expression without binding variable referral.
802
If all the sub-expressions have negation (or there is just one sub-expression and it has negation) then augment the expression with sub-expression that is the corresponding output-dimension, without any additional constraint. Then redo the reordering as above.
803
If there is referral to binding variables in the sub-expression then use the IN structure approach in the SQL translation, otherwise use MINUS structure.

FIG. 8

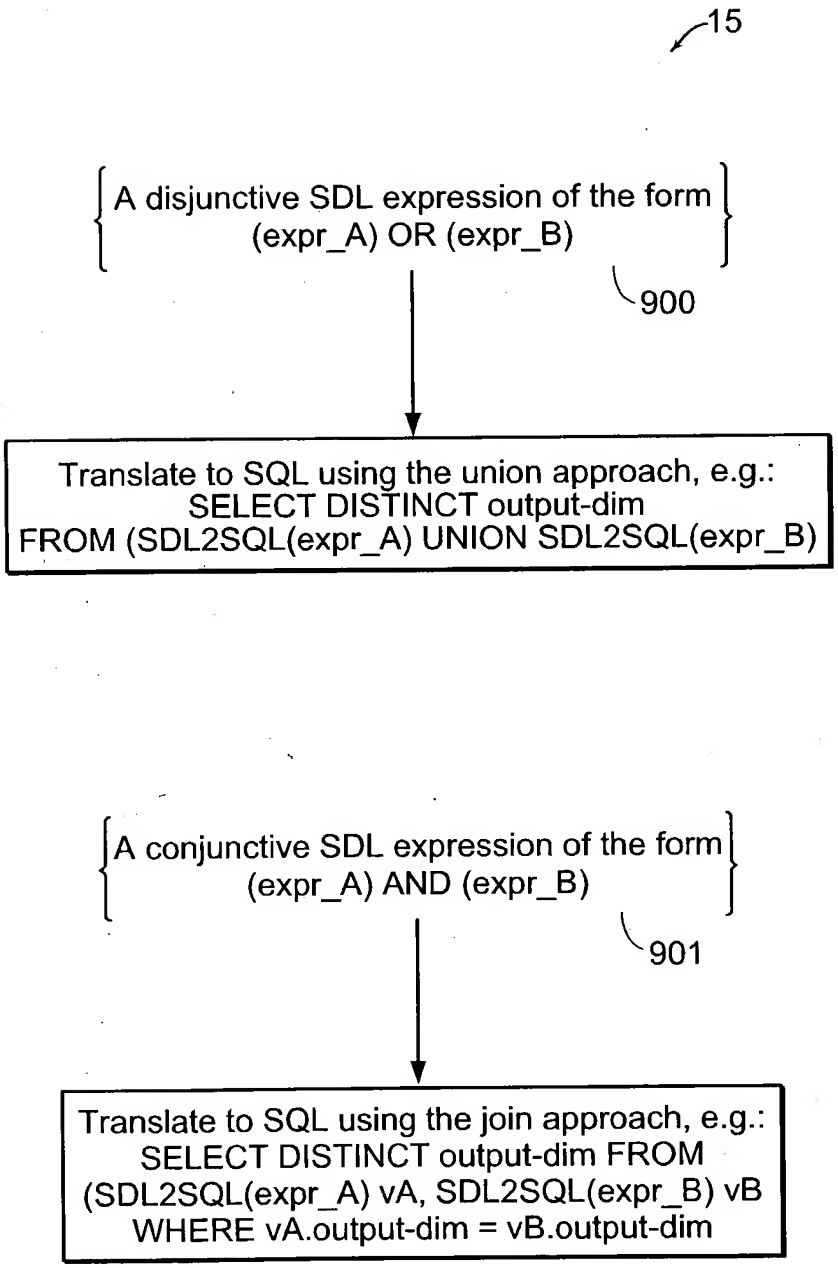
15
A disjunctive SDL expression of the form (expr_A) OR (expr_B)
900
Translate to SQL using the union approach, e.g.: SELECT DISTINCT output-dim FROM (SDL2SQL(expr_A) UNION SDL2SQL(expr_B)
A conjunctive SDL expression of the form (expr_A) AND (expr_B)
901
Translate to SQL using the join approach, e.g.: SELECT DISTINCT output-dim FROM (SDL2SQL(expr_A) vA, SDL2SQL(expr_B) vB WHERE vA.output-dim = vB.output-dim

FIG. 9

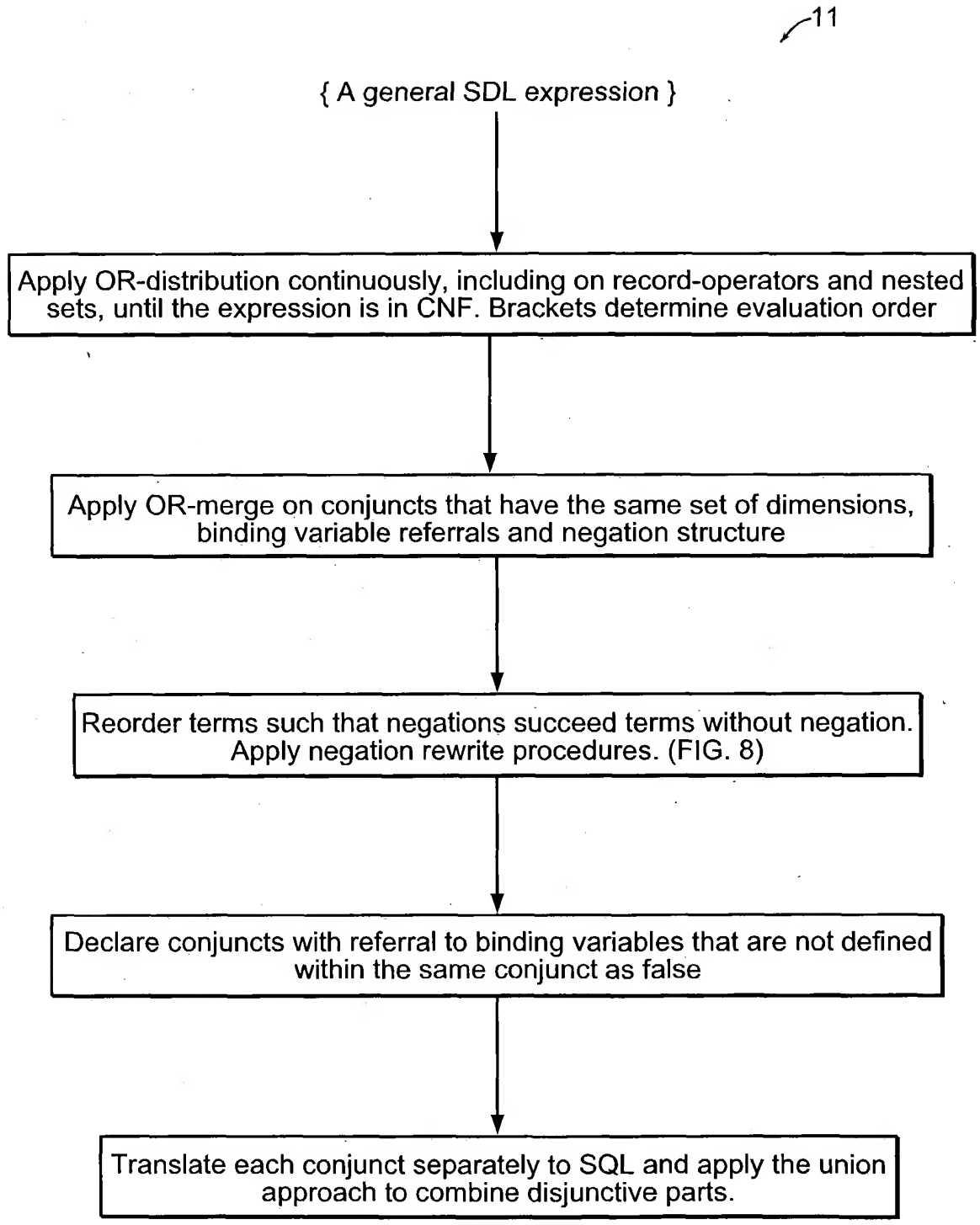
11
{ A general SDL expression }
Apply OR-distribution continuously, including on record-operators and nested sets, until the expression is in CNF. Brackets determine evaluation order
Apply OR-merge on conjuncts that have the same set of dimensions, binding variable referrals and negation structure
Reorder terms such that negations succeed terms without negation. Apply negation rewrite procedures. (FIG. 8)
Declare conjuncts with referral to binding variables that are not defined within the same conjunct as false
Translate each conjunct separately to SQL and apply the union approach to combine disjunctive parts.

FIG. 10

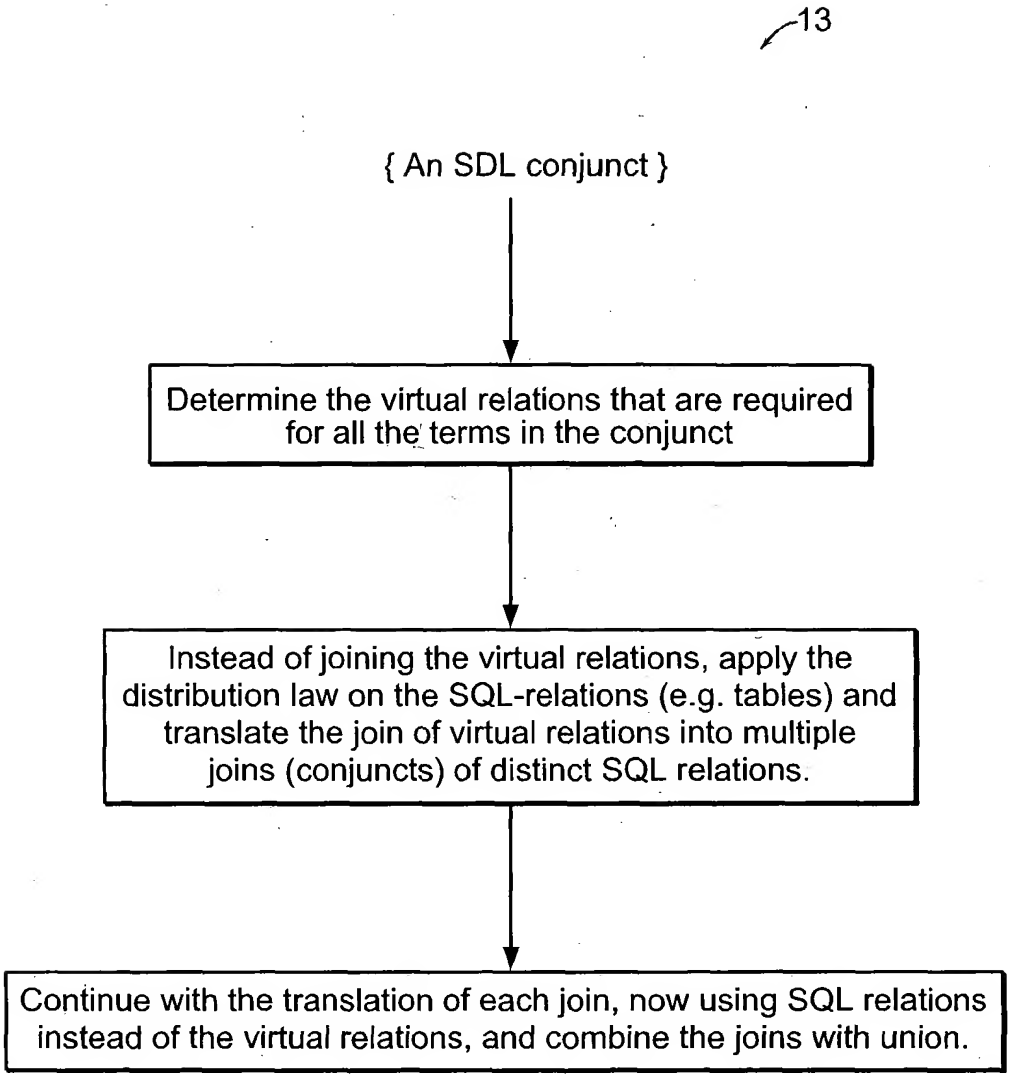
13
{ An SDL conjunct }
Determine the virtual relations that are required for all the terms in the conjunct
Instead of joining the virtual relations, apply the distribution law on the SQL-relations (e.g. tables) and translate the join of virtual relations into multiple joins (conjuncts) of distinct SQL relations.
Continue with the translation of each join, now using SQL relations instead of the virtual relations, and combine the joins with union.

FIG. 11

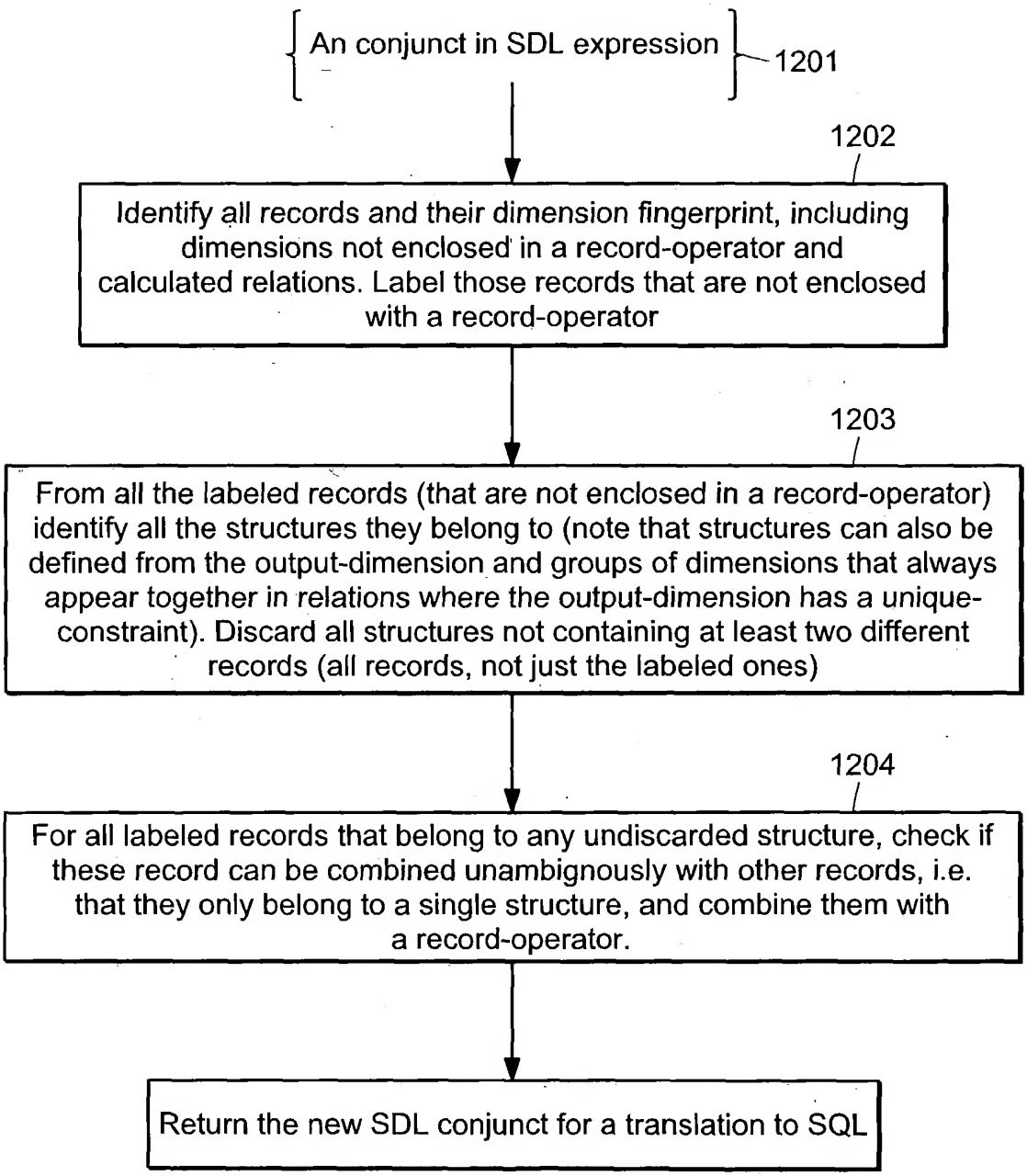
An conjunct in SDL expression
1201
1202
Identify all records and their dimension fingerprint, including dimensions not enclosed in a record-operator and calculated relations. Label those records that are not enclosed with a record-operator
1203
From all the labeled records (that are not enclosed in a record-operator) identify all the structures they belong to (note that structures can also be defined from the output-dimension and groups of dimensions that always appear together in relations where the output-dimension has a unique-constraint). Discard all structures not containing at least two different records (all records, not just the labeled ones)
1204
For all labeled records that belong to any undiscarded structure, check if these record can be combined unambignously with other records, i.e. that they only belong to a single structure, and combine them with a record-operator.
Return the new SDL conjunct for a translation to SQL

FIG. 12

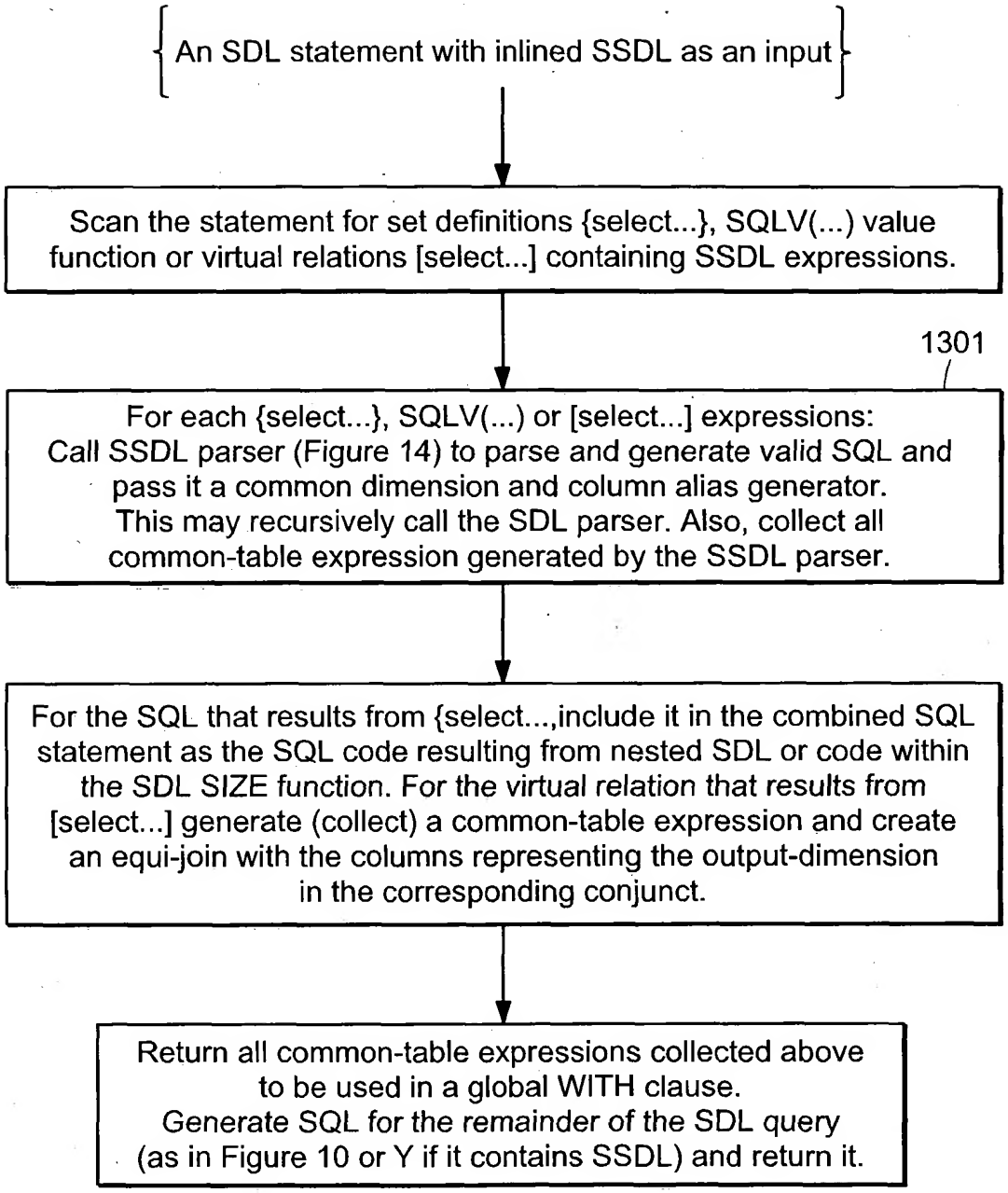
An SDL statement with inlined SSDL as an input
Scan the statement for set definitions {select...}, SQLV(...) value function or virtual relations [select...] containing SSDL expressions.
1301
For each {select...}, SQLV(...) or [select...] expressions:
Call SSDL parser (Figure 14) to parse and generate valid SQL and pass it a common dimension and column alias generator.
This may recursively call the SDL parser. Also, collect all common-table expression generated by the SSDL parser.
For the SQL that results from {select...,include it in the combined SQL statement as the SQL code resulting from nested SDL or code within the SDL SIZE function. For the virtual relation that results from [select...] generate (collect) a common-table expression and create an equi-join with the columns representing the output-dimension in the corresponding conjunct.
Return all common-table expressions collected above to be used in a global WITH clause.
Generate SQL for the remainder of the SDL query (as in Figure 10 or Y if it contains SSDL) and return it.

FIG. 13

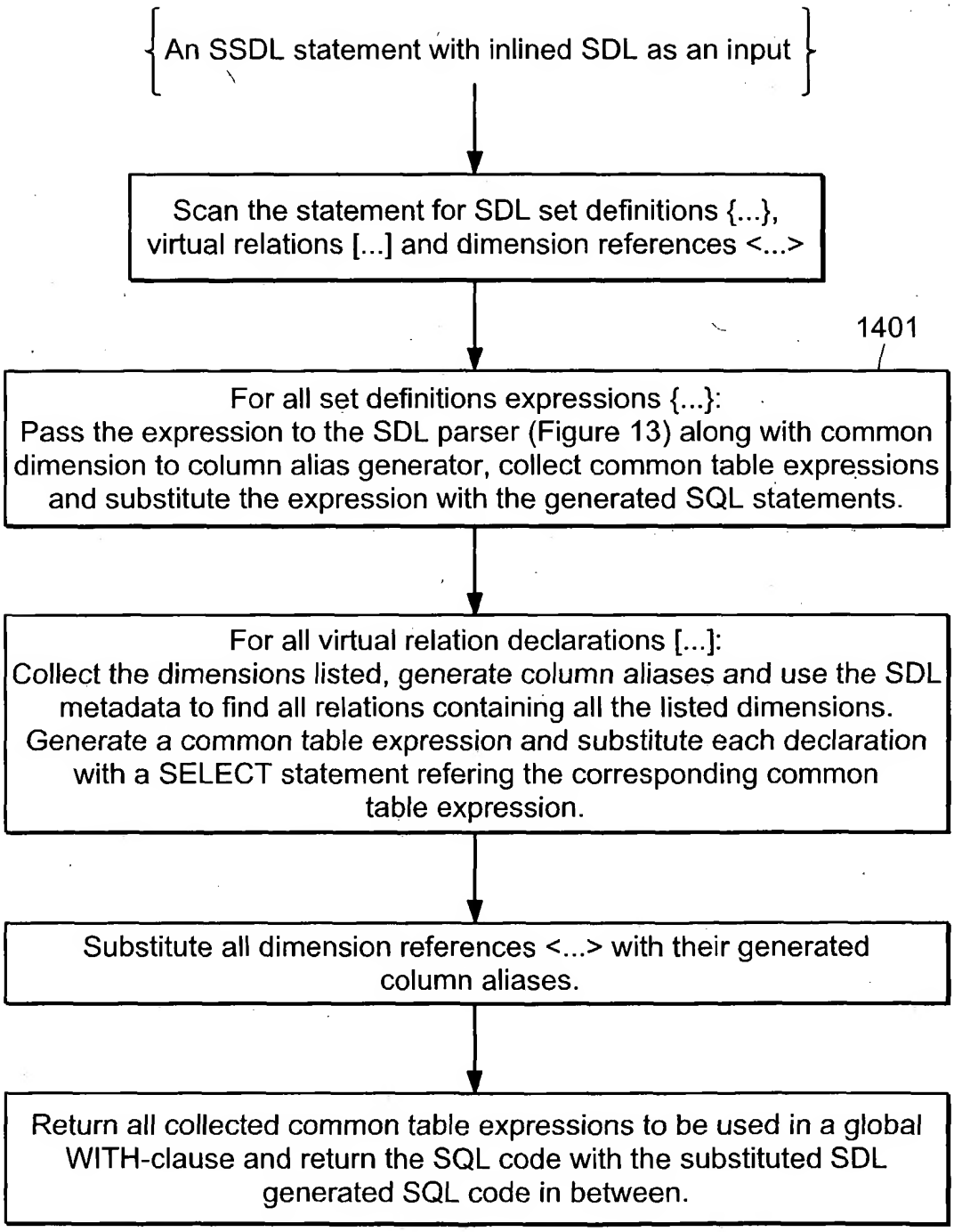
An SSDL statement with inlined SDL as an input
Scan the statement for SDL set definitions {...}, virtual relations [...] and dimension references <...>
1401
For all set definitions expressions {...}:
Pass the expression to the SDL parser (Figure 13) along with common dimension to column alias generator, collect common table expressions and substitute the expression with the generated SQL statements.
For all virtual relation declarations [...]:
Collect the dimensions listed, generate column aliases and use the SDL metadata to find all relations containing all the listed dimensions.
Generate a common table expression and substitute each declaration with a SELECT statement refering the corresponding common table expression.
Substitute all dimension references <...> with their generated column aliases.
Return all collected common table expressions to be used in a global WITH-clause and return the SQL code with the substituted SDL generated SQL code in between.

FIG. 14

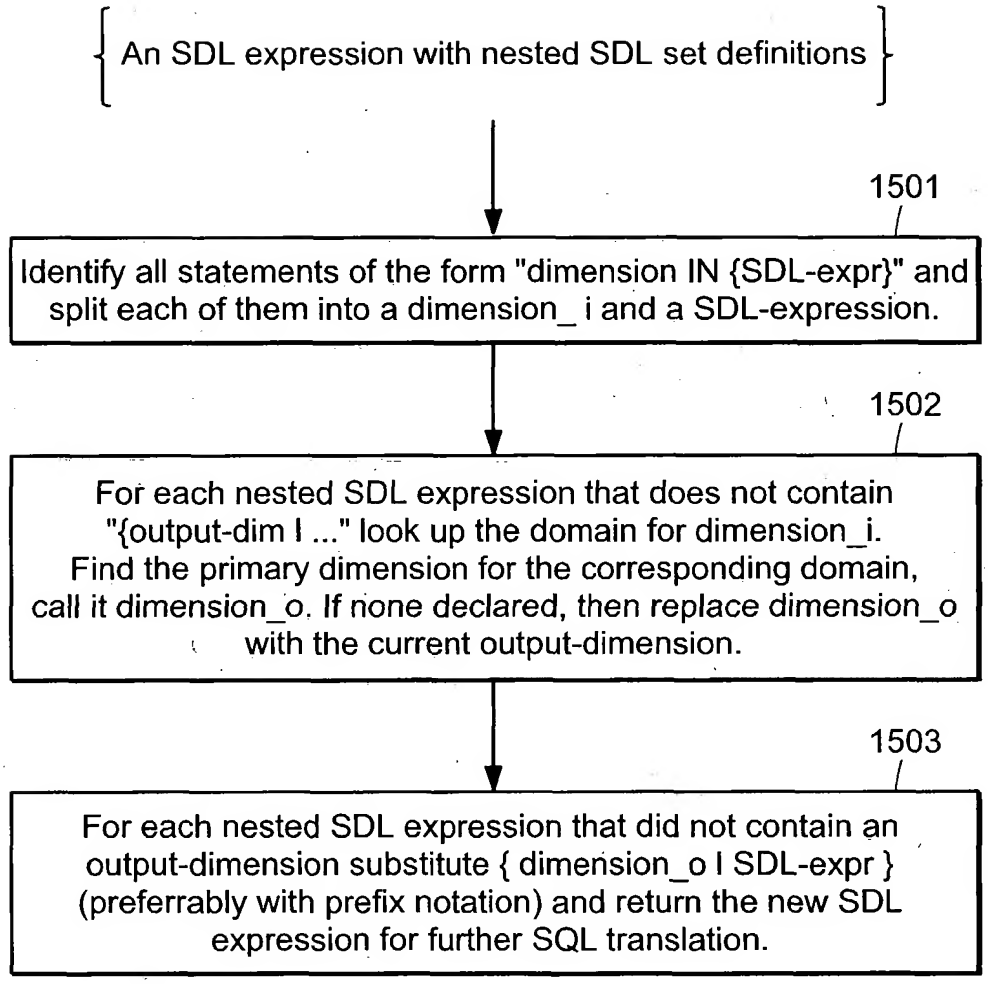
An SDL expression with nested SDL set definitions
1501
Identify all statements of the form "dimension IN {SDL-expr}" and split each of them into a dimension_ i and a SDL-expression.
1502
For each nested SDL expression that does not contain "{output-dim I ..." look up the domain for dimension_i. Find the primary dimension for the corresponding domain, call it dimension_o. If none declared, then replace dimension_o with the current output-dimension.
1503
For each nested SDL expression that did not contain an output-dimension substitute { dimension_o I SDL-expr } (preferrably with prefix notation) and return the new SDL expression for further SQL translation.

FIG. 15

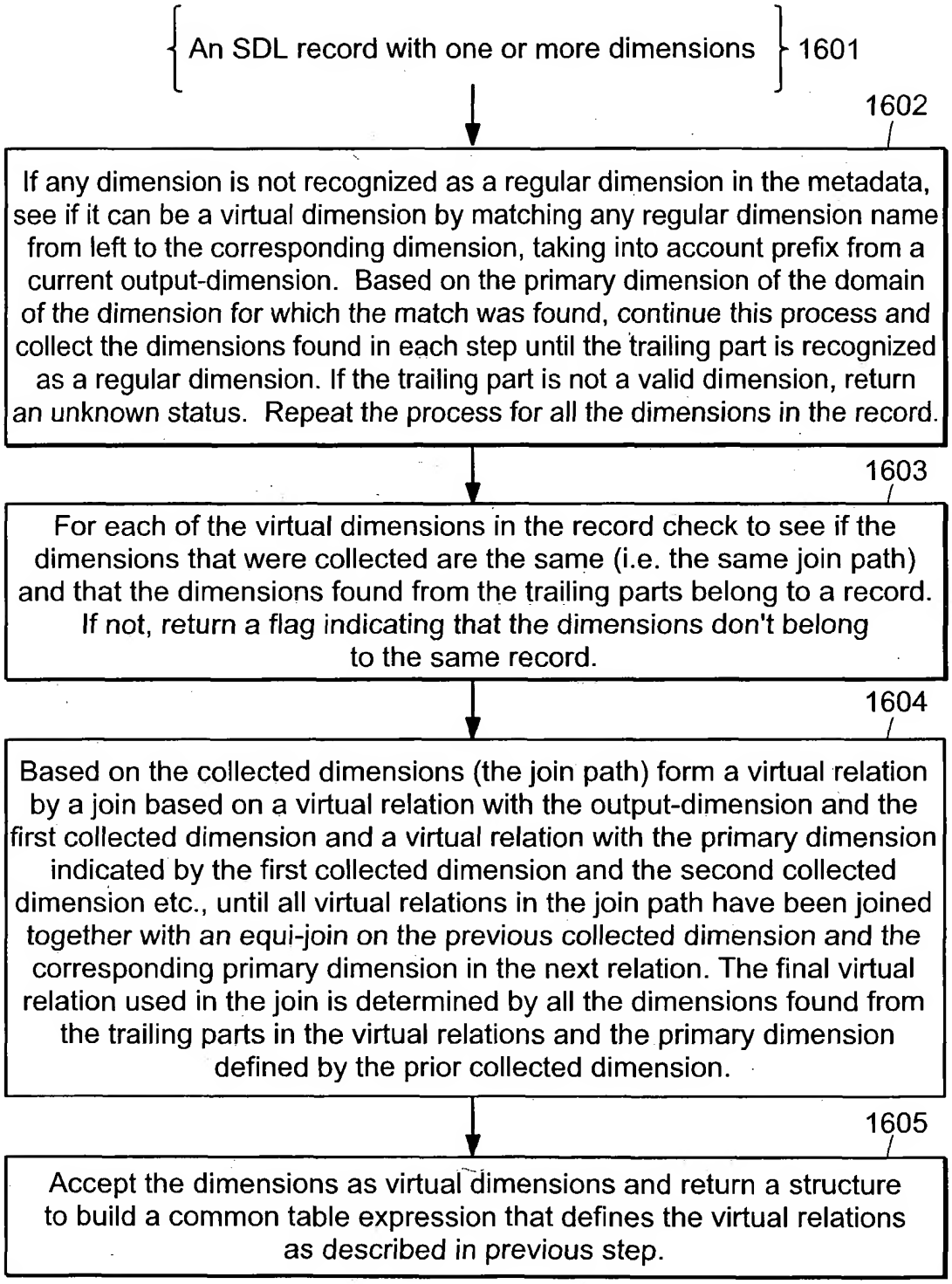
An SDL record with one or more dimensions
1601
1602
If any dimension is not recognized as a regular dimension in the metadata, see if it can be a virtual dimension by matching any regular dimension name from left to the corresponding dimension, taking into account prefix from a current output-dimension. Based on the primary dimension of the domain of the dimension for which the match was found, continue this process and collect the dimensions found in each step until the trailing part is recognized as a regular dimension. If the trailing part is not a valid dimension, return an unknown status. Repeat the process for all the dimensions in the record.
1603
For each of the virtual dimensions in the record check to see if the dimensions that were collected are the same (i.e. the same join path) and that the dimensions found from the trailing parts belong to a record. If not, return a flag indicating that the dimensions don't belong to the same record.
1604
Based on the collected dimensions (the join path) form a virtual relation by a join based on a virtual relation with the output-dimension and the first collected dimension and a virtual relation with the primary dimension indicated by the first collected dimension and the second collected dimension etc., until all virtual relations in the join path have been joined together with an equi-join on the previous collected dimension and the corresponding primary dimension in the next relation. The final virtual relation used in the join is determined by all the dimensions found from the trailing parts in the virtual relations and the primary dimension defined by the prior collected dimension.
1605
Accept the dimensions as virtual dimensions and return a structure to build a common table expression that defines the virtual relations as described in previous step.

FIG. 16

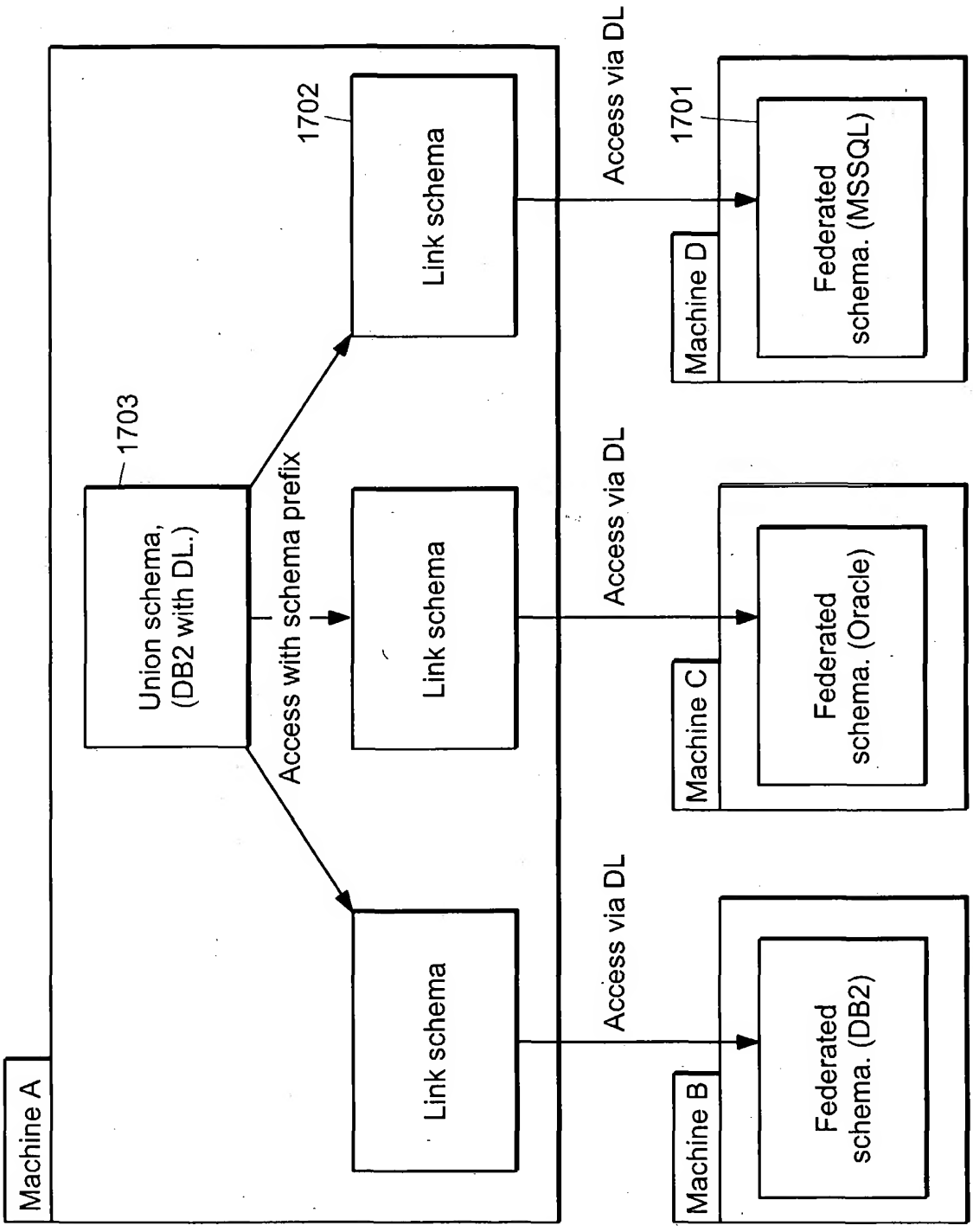
Machine A
Union schema, (DB2 with DL.)
1703
Access with schema prefix
Link schema
1702
Link schema
Link schema
Access via DL
Access via DL
Access via DL
Machine B
Federated schema. (DB2)
Machine C
Federated schema. (Oracle)
Machine D
Federated schema. (MSSQL)
1701

FIG. 17

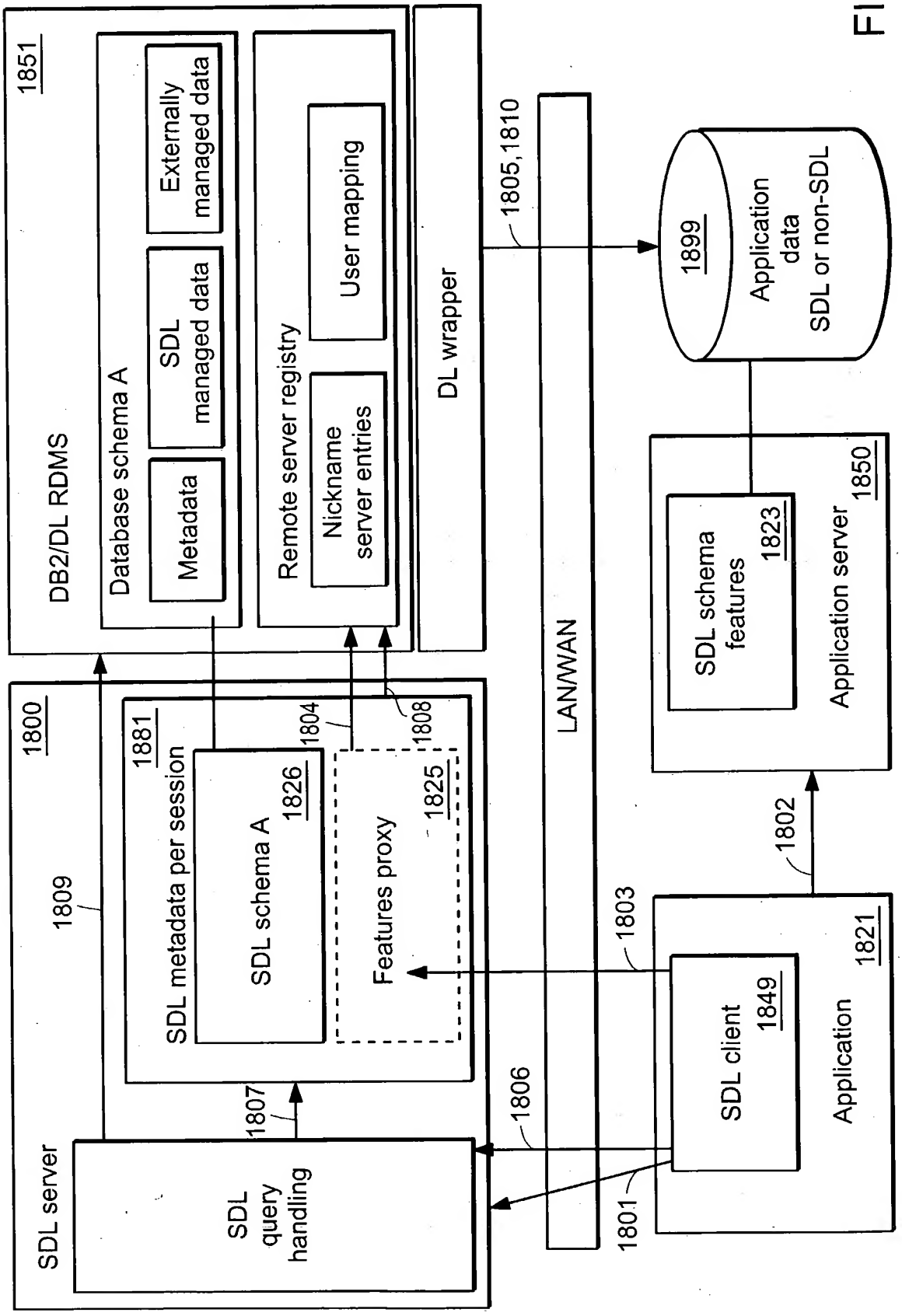
1851
DB2/DL RDMS
Database schema A
Metadata
SDL managed data
Externally managed data
Remote server registry
Nickname server entries
User mapping
DL wrapper
1805,1810
1899
Application data
SDL or non-SDL
LAN/WAN
1800
1809
1881
SDL metadata per session
SDL schema A
1826
1804
1808
Features proxy
1825
SDL schema features
1823
Application server
1850
1802
1803
SDL client
1849
Application
1821
1807
1806
SDL server
SDL query handling
1801

FIG. 18

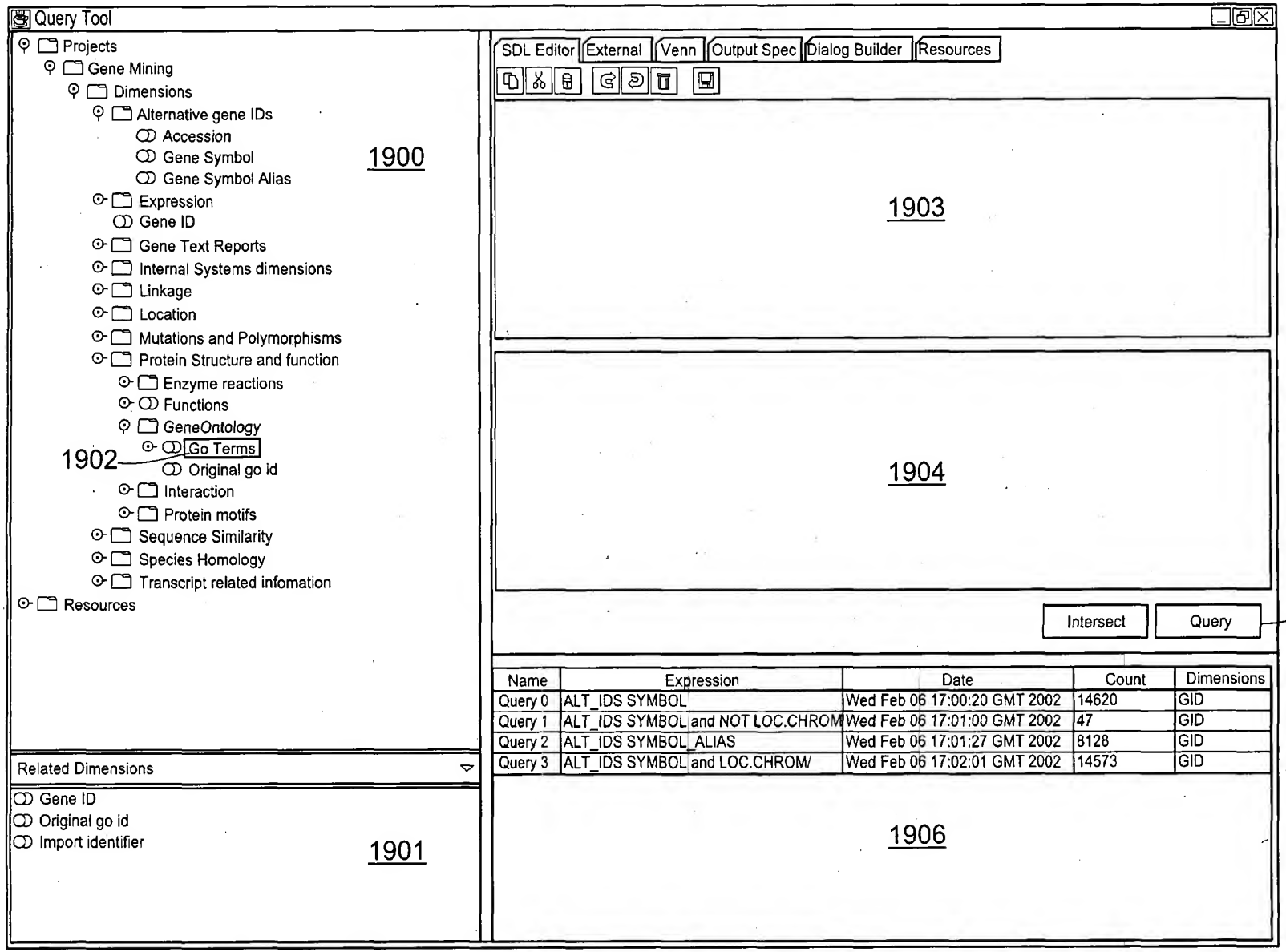

FIG. 19A

FIG. 19B

FIG. 19

FIG. 19A

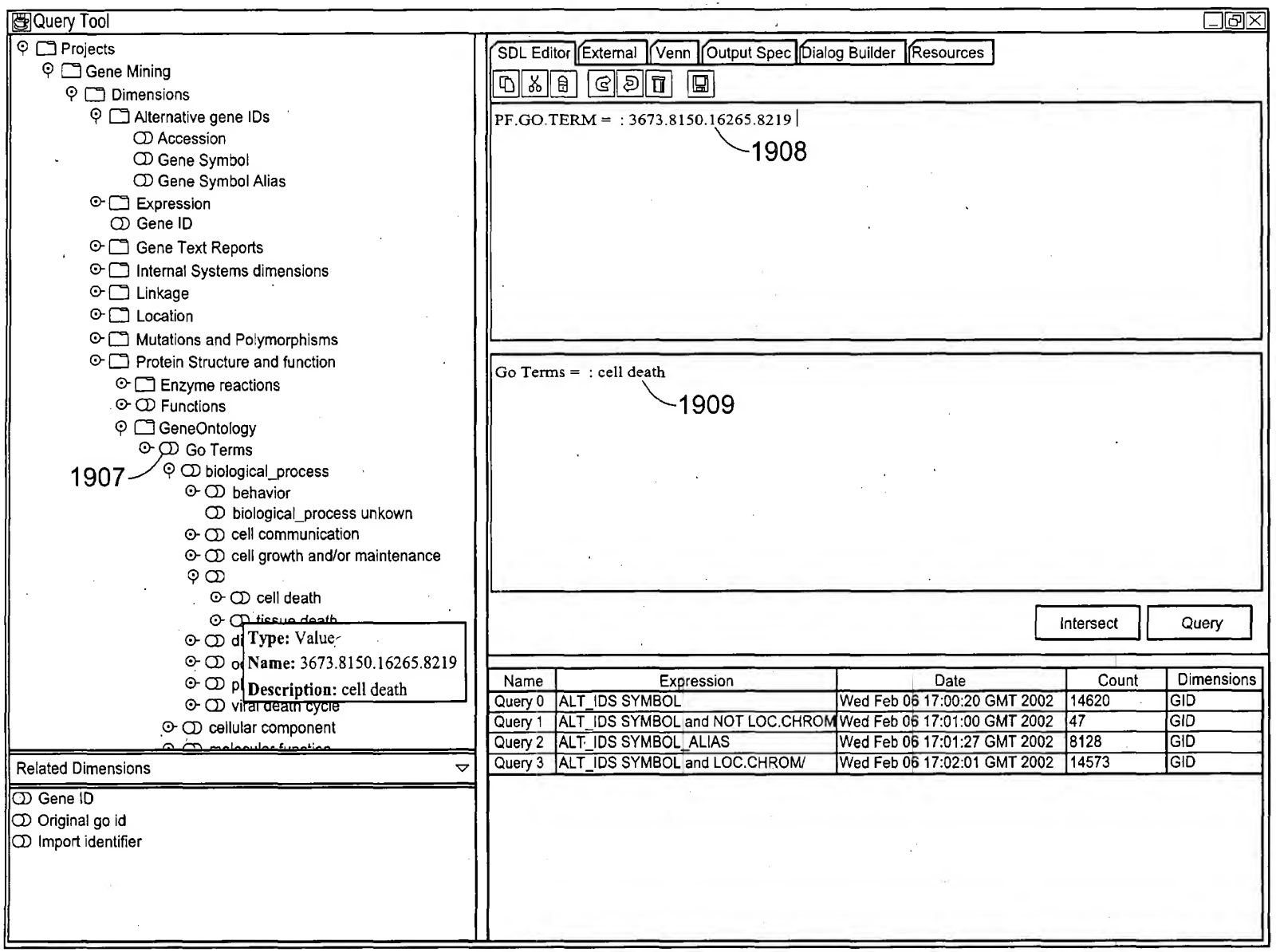
Query Tool
Projects
Gene Mining
Dimensions
Alternative gene IDs
Accession
Gene Symbol
Gene Symbol Alias
Expression
Gene ID
Gene Text Reports
Internal Systems dimensions
Linkage
Location
Mutations and Polymorphisms
Protein Structure and function
Enzyme reactions
Functions
GeneOntology
Go Terms
1907
biological_process
behavior
biological_process unkown
cell communication
cell growth and/or maintenance
cell death
cellular component
Type: Value
Name: 3673.8150.16265.8219
Description: cell death
Related Dimensions
Gene ID
Original go id
Import identifier
SDL Editor
External
Venn
Output Spec
Dialog Builder
Resources
PF.GO.TERM = : 3673.8150.16265.8219
1908
Go Terms = : cell death
1909
Intersect
Query
Name
Expression
Date
Count
Dimensions
Query 0
ALT_IDS SYMBOL
Wed Feb 06 17:00:20 GMT 2002
14620
GID
Query 1
ALT_IDS SYMBOL and NOT LOC.CHROM
Wed Feb 06 17:01:00 GMT 2002
47
GID
Query 2
ALT_IDS SYMBOL_ALIAS
Wed Feb 06 17:01:27 GMT 2002
8128
GID
Query 3
ALT_IDS SYMBOL and LOC.CHROM/
Wed Feb 06 17:02:01 GMT 2002
14573
GID

FIG. 19B

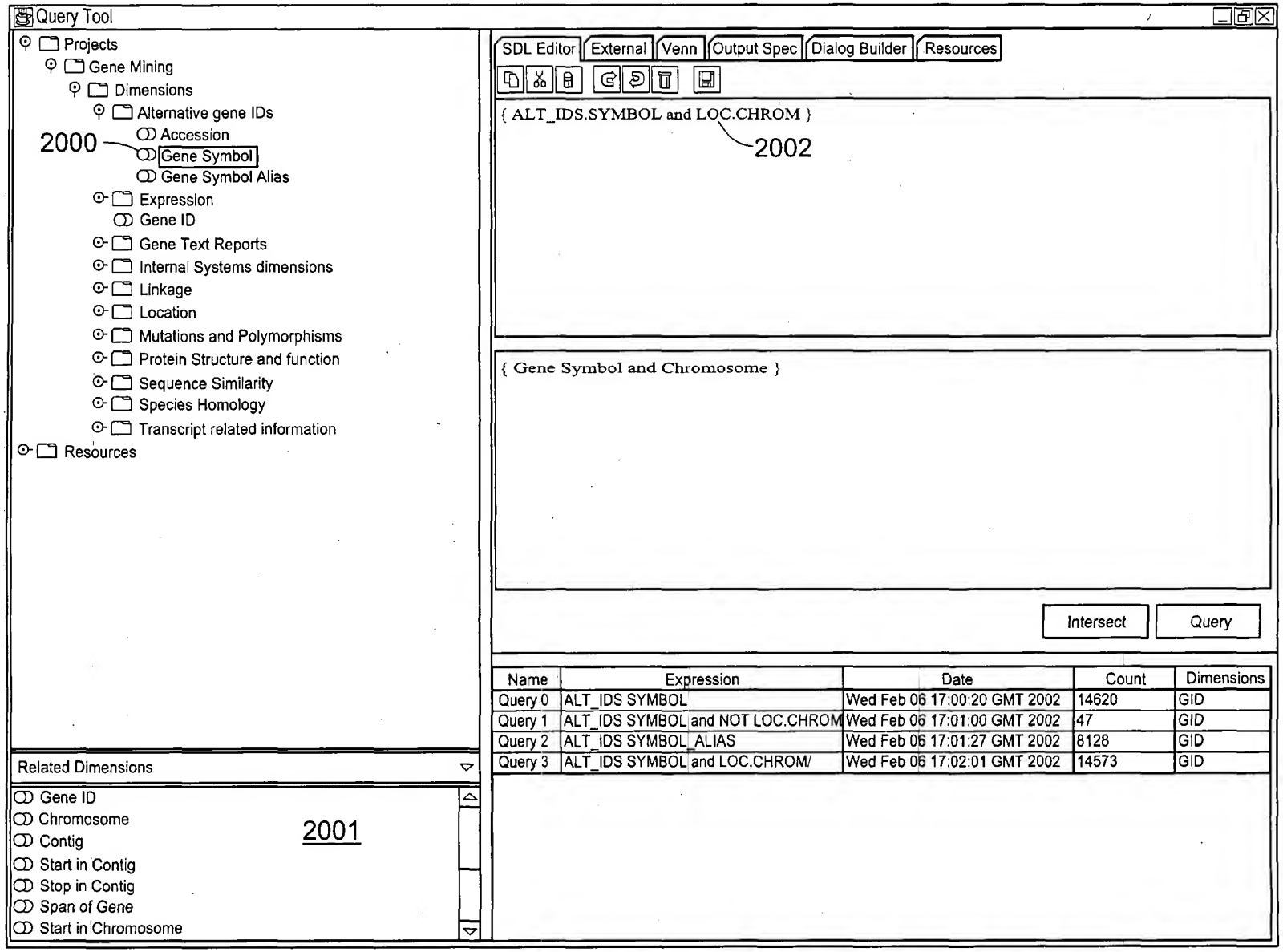

FIG. 20A

FIG. 20B

FIG. 20

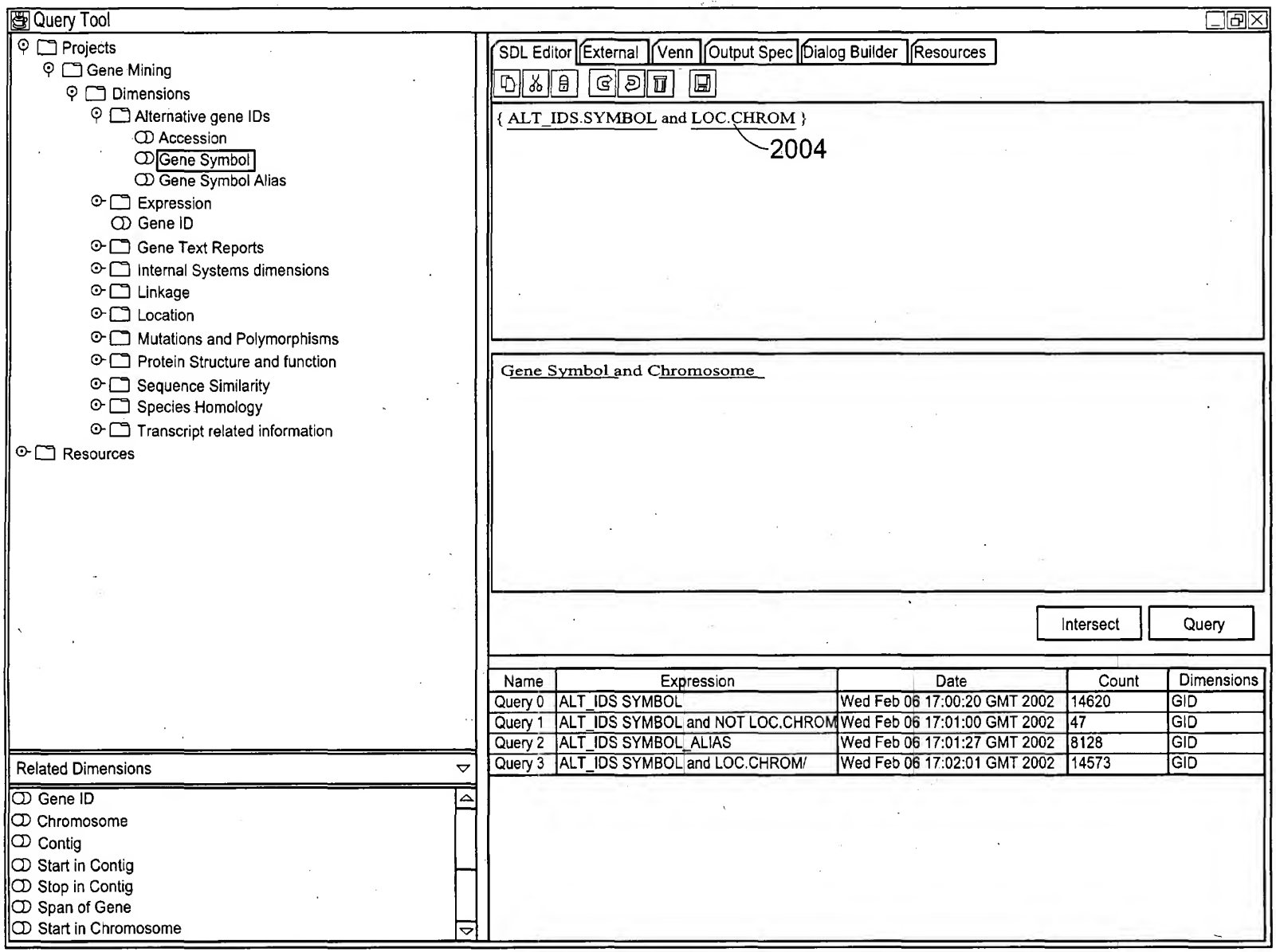
Query Tool
Projects
Gene Mining
Dimensions
Alternative gene IDs
Accession
Gene Symbol
Gene Symbol Alias
Expression
Gene ID
Gene Text Reports
Internal Systems dimensions
Linkage
Location
Mutations and Polymorphisms
Protein Structure and function
Sequence Similarity
Species Homology
Transcript related information
Resources
SDL Editor
External
Venn
Output Spec
Dialog Builder
Resources
{ ALT_IDS.SYMBOL and LOC.CHROM }
2004
Gene Symbol and Chromosome
Intersect
Query
Name
Expression
Date
Count
Dimensions
Query 0
ALT_IDS SYMBOL
Wed Feb 06 17:00:20 GMT 2002
14620
GID
Query 1
ALT_IDS SYMBOL and NOT LOC.CHROM
Wed Feb 06 17:01:00 GMT 2002
47
GID
Query 2
ALT_IDS SYMBOL_ALIAS
Wed Feb 06 17:01:27 GMT 2002
8128
GID
Query 3
ALT_IDS SYMBOL and LOC.CHROM/
Wed Feb 06 17:02:01 GMT 2002
14573
GID
Related Dimensions
Gene ID
Chromosome
Contig
Start in Contig
Stop in Contig
Span of Gene
Start in Chromosome

FIG. 20B

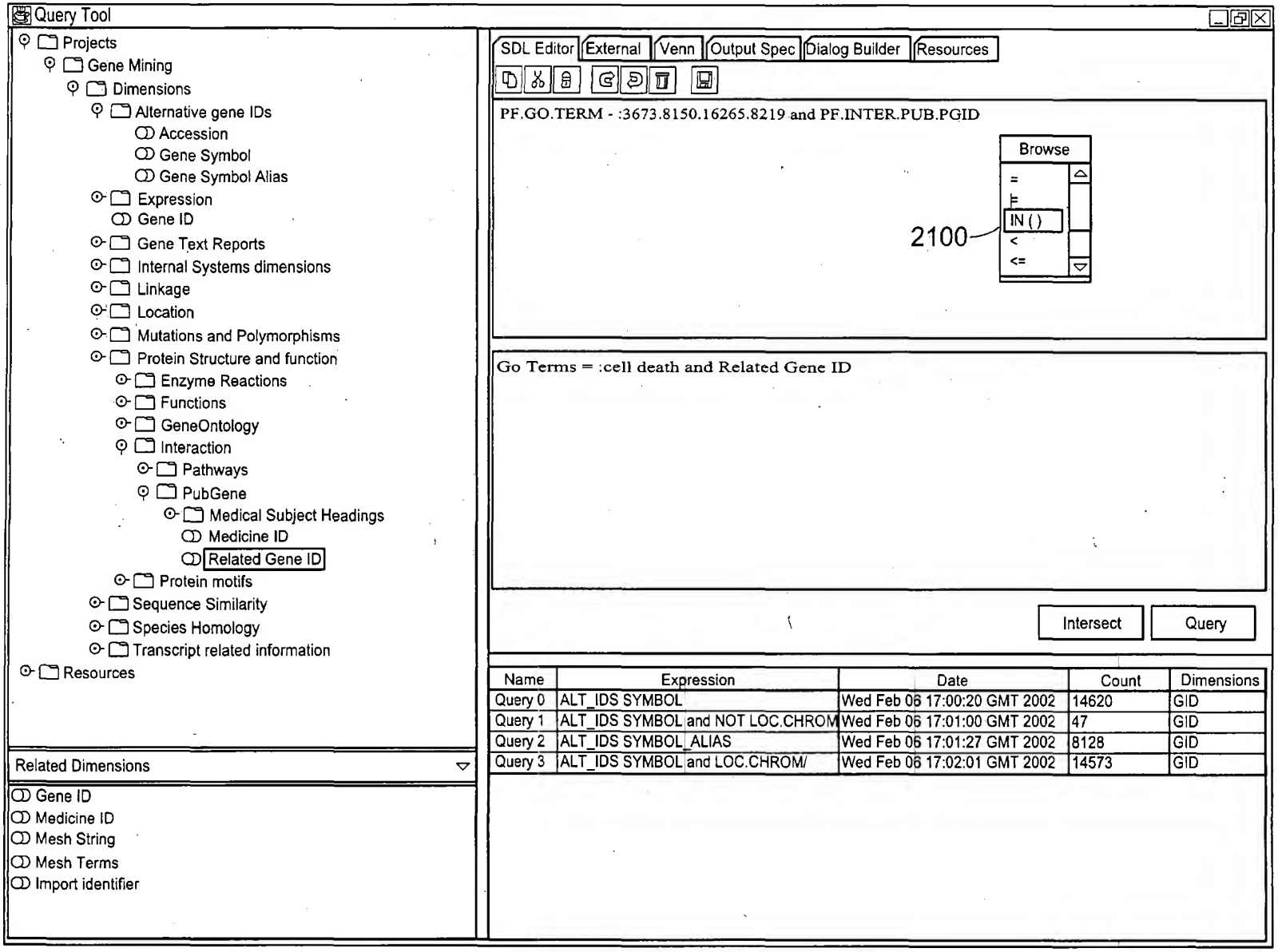

FIG. 21A

FIG. 21B

FIG. 21

FIG. 21A

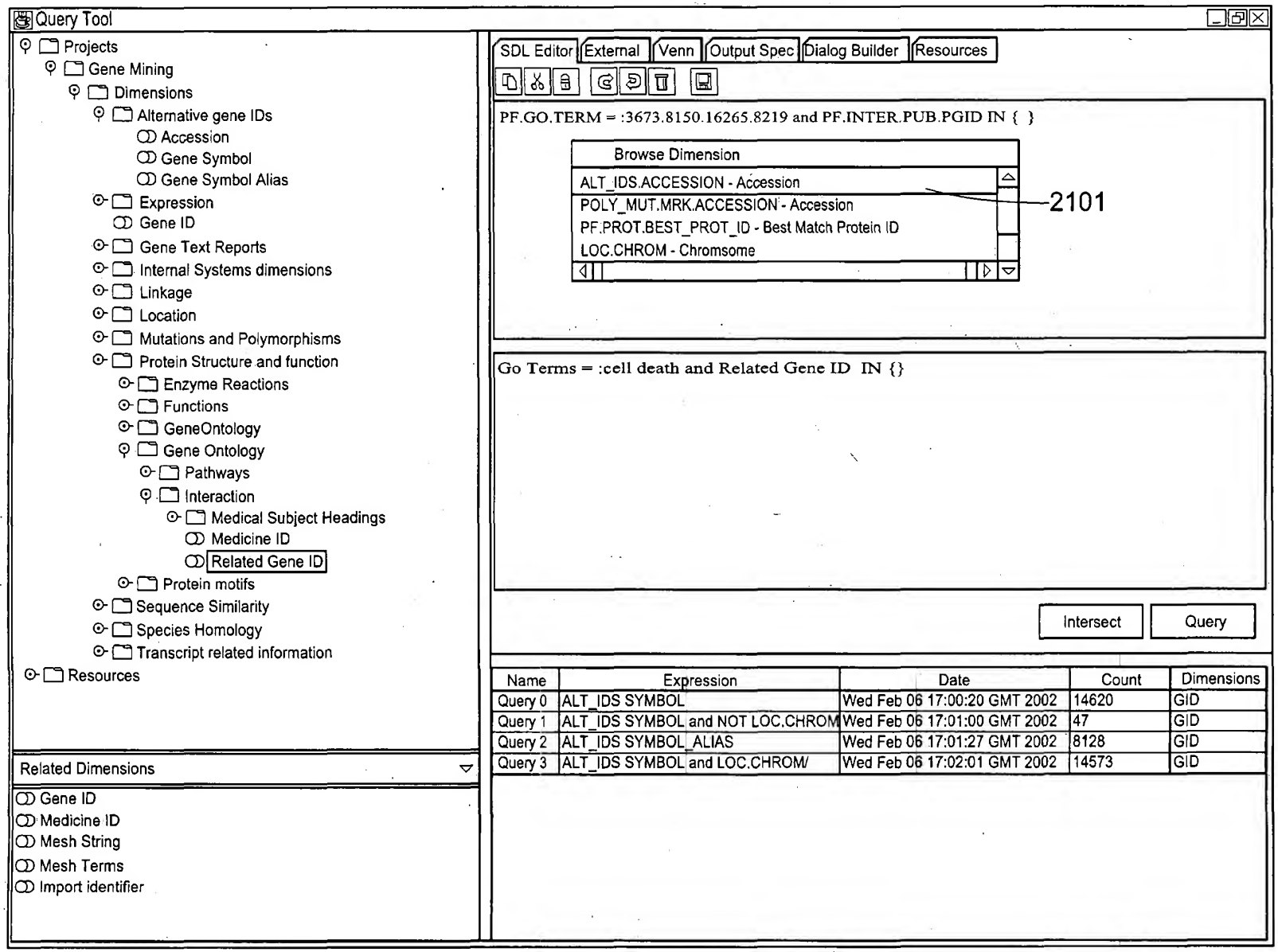
Query Tool
Projects
Gene Mining
Dimensions
Alternative gene IDs
Accession
Gene Symbol
Gene Symbol Alias
Expression
Gene ID
Gene Text Reports
Internal Systems dimensions
Linkage
Location
Mutations and Polymorphisms
Protein Structure and function
Enzyme Reactions
Functions
GeneOntology
Gene Ontology
Pathways
Interaction
Medical Subject Headings
Medicine ID
Related Gene ID
Protein motifs
Sequence Similarity
Species Homology
Transcript related information
Resources
Related Dimensions
Gene ID
Medicine ID
Mesh String
Mesh Terms
Import identifier
SDL Editor
External
Venn
Output Spec
Dialog Builder
Resources
PF.GO.TERM = :3673.8150.16265.8219 and PF.INTER.PUB.PGID IN { }
Browse Dimension
ALT_IDS.ACCESSION - Accession
POLY_MUT.MRK.ACCESSION - Accession
PF.PROT.BEST_PROT_ID - Best Match Protein ID
LOC.CHROM - Chromsome
2101
Go Terms = :cell death and Related Gene ID IN {}
Intersect
Query
Name
Expression
Date
Count
Dimensions
Query 0
ALT_IDS SYMBOL
Wed Feb 06 17:00:20 GMT 2002
14620
GID
Query 1
ALT_IDS SYMBOL and NOT LOC.CHROM
Wed Feb 06 17:01:00 GMT 2002
47
GID
Query 2
ALT_IDS SYMBOL_ALIAS
Wed Feb 06 17:01:27 GMT 2002
8128
GID
Query 3
ALT_IDS SYMBOL and LOC.CHROM/
Wed Feb 06 17:02:01 GMT 2002
14573
GID

FIG. 21B

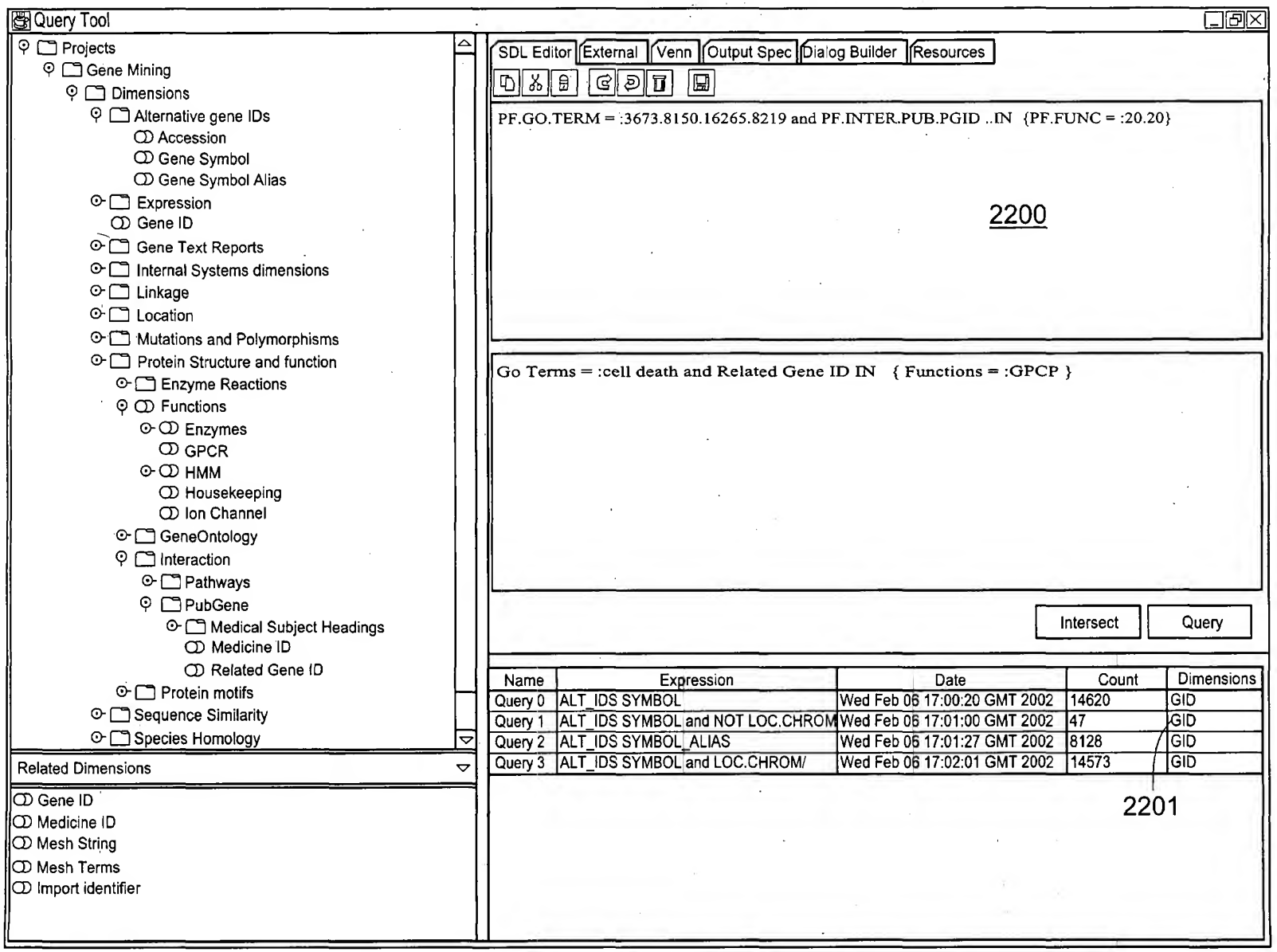

FIG. 22A

FIG. 22B

FIG. 22

FIG. 22A

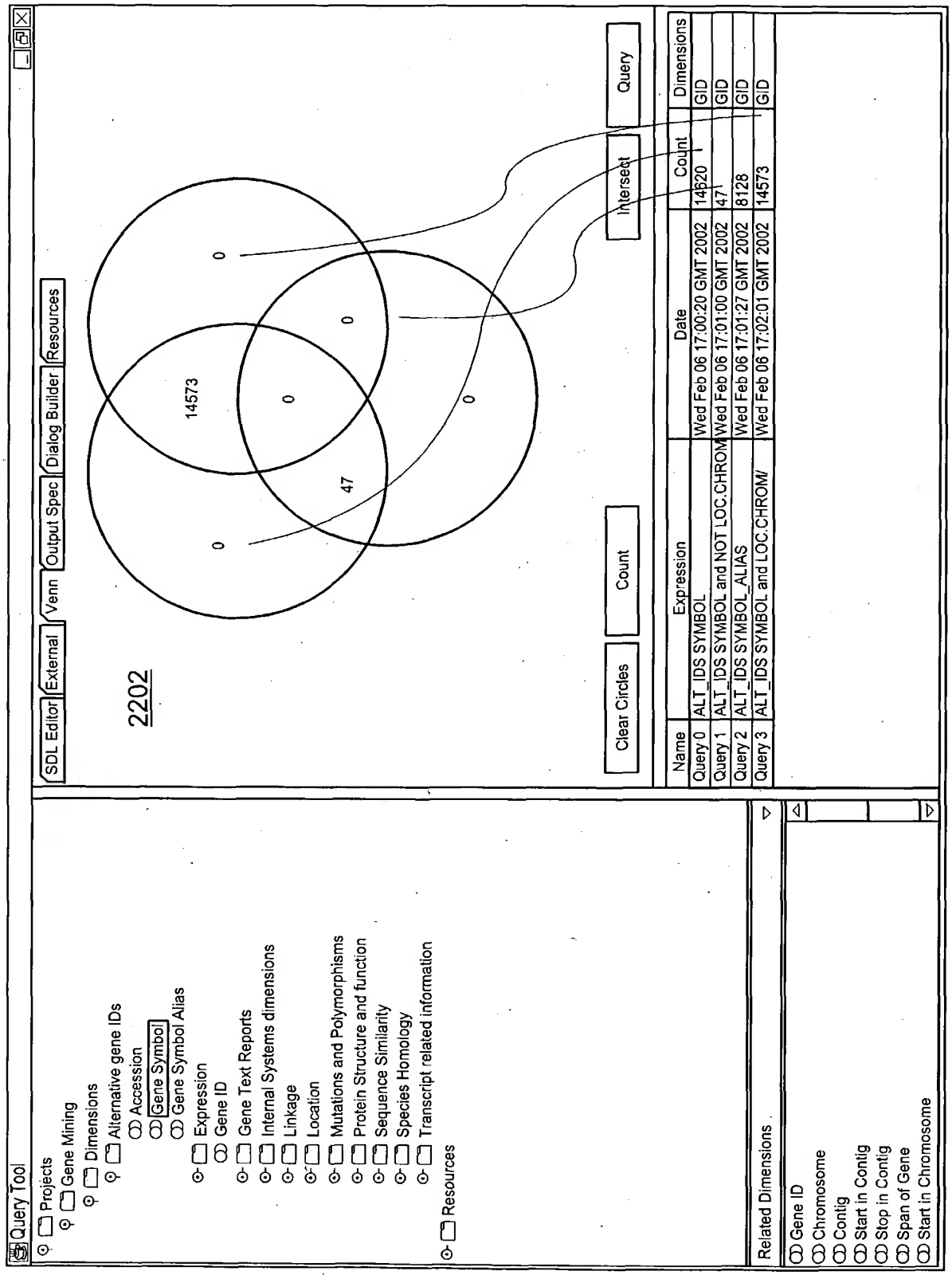
Query Tool
Projects
Gene Mining
Dimensions
Alternative gene IDs
Accession
Gene Symbol
Gene Symbol Alias
Expression
Gene ID
Gene Text Reports
Internal Systems dimensions
Linkage
Location
Mutations and Polymorphisms
Protein Structure and function
Sequence Similarity
Species Homology
Transcript related information
Resources
Related Dimensions
Gene ID
Chromosome
Contig
Start in Contig
Stop in Contig
Span of Gene
Start in Chromosome
SDL Editor
External
Venn
Output Spec
Dialog Builder
Resources
2202
14573
0
0
0
0
0
47
0
Clear Circles
Count
Intersect
Query
Name
Expression
Date
Count
Dimensions
Query 0
ALT_IDS SYMBOL
Wed Feb 06 17:00:20 GMT 2002
14620
GID
Query 1
ALT_IDS SYMBOL and NOT LOC.CHROM
Wed Feb 06 17:01:00 GMT 2002
47
GID
Query 2
ALT_IDS SYMBOL_ALIAS
Wed Feb 06 17:01:27 GMT 2002
8128
GID
Query 3
ALT_IDS SYMBOL and LOC.CHROM/
Wed Feb 06 17:02:01 GMT 2002
14573
GID

FIG. 22B

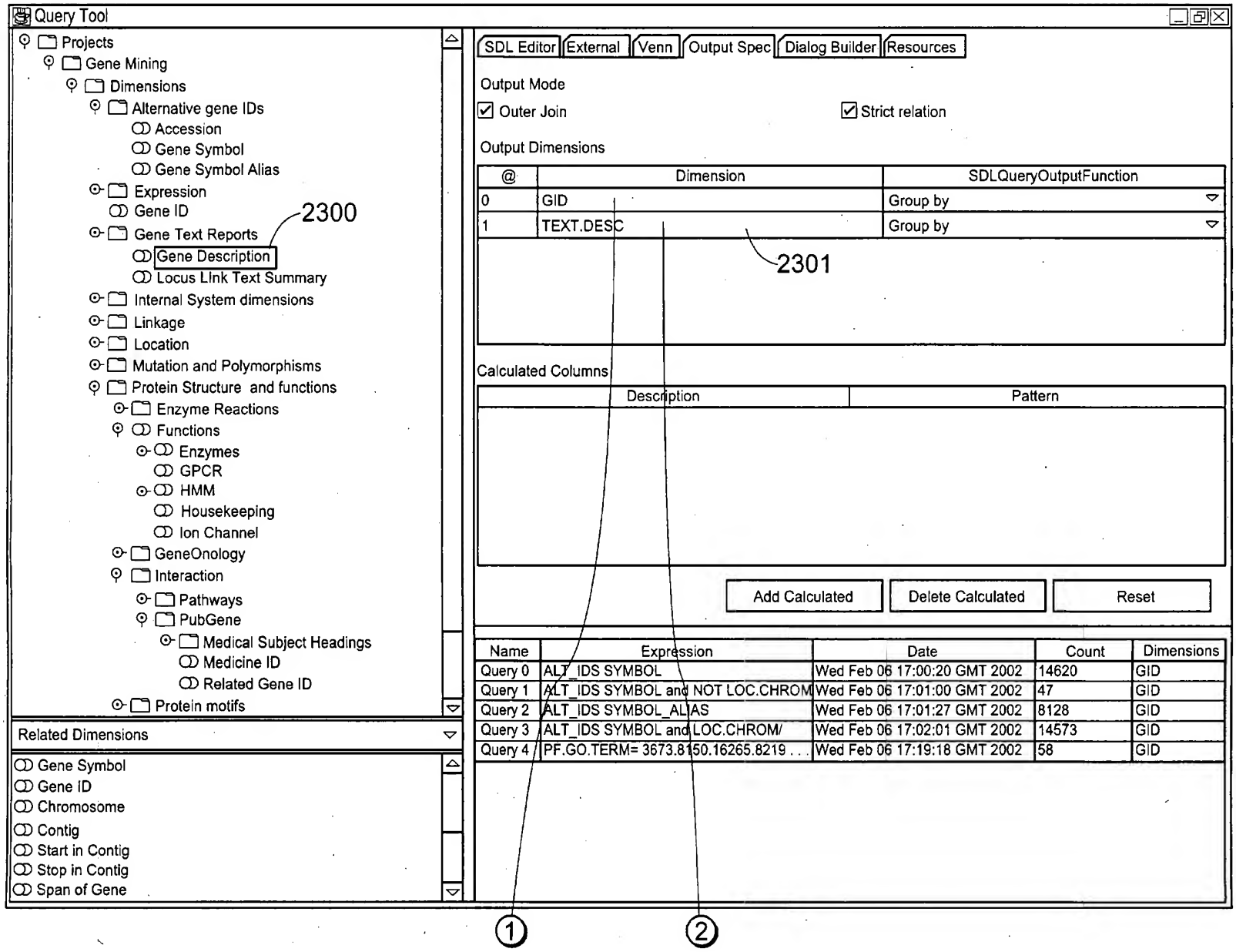

FIG. 23A

FIG. 23B

FIG. 23

FIG. 23A

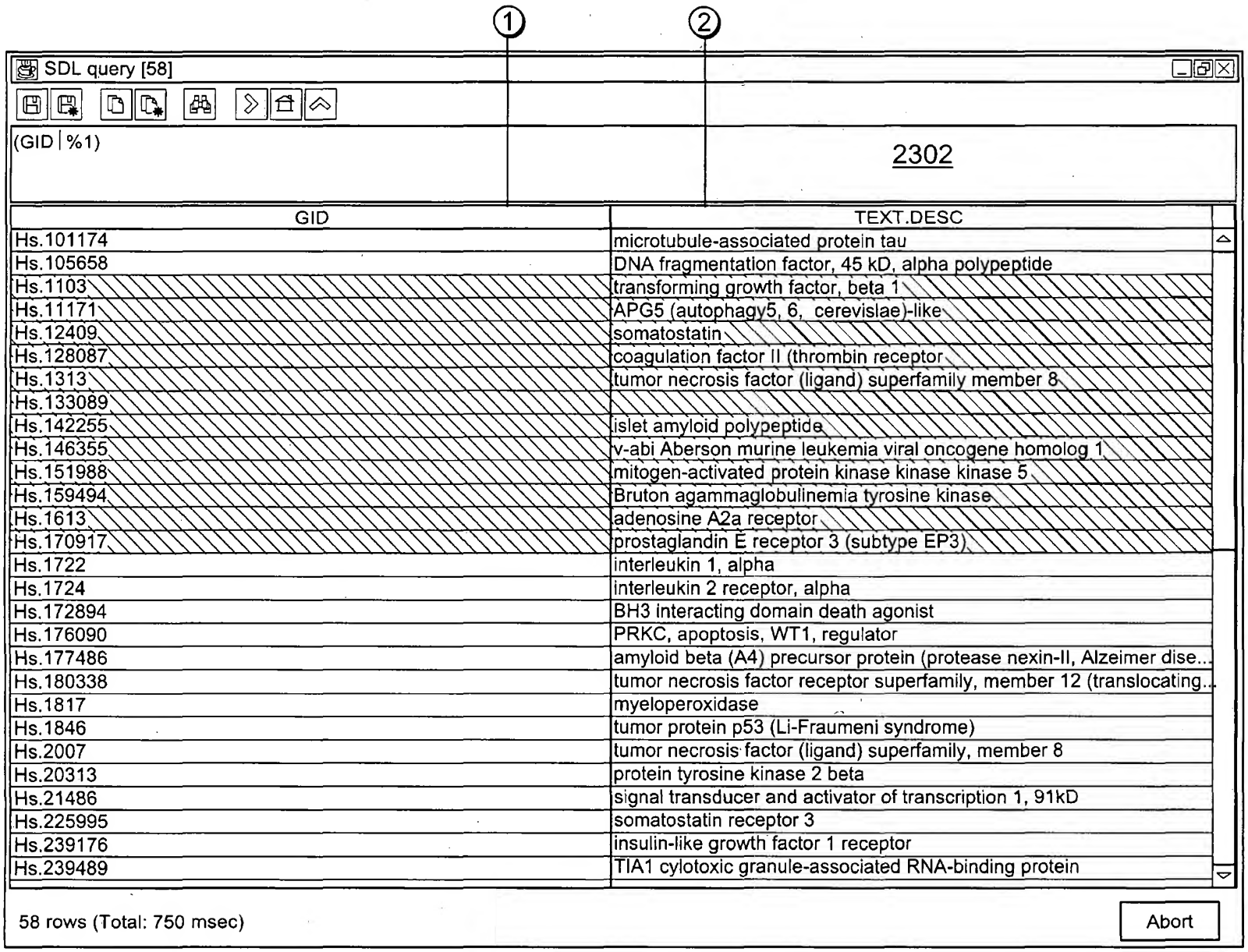
1
2
SDL query [58]
(GID | %1)
2302
GID
TEXT.DESC
Hs.101174 microtubule-associated protein tau
Hs.105658 DNA fragmentation factor, 45 kD, alpha polypeptide
Hs.1103 transforming growth factor, beta 1
Hs.11171 APG5 (autophagy5, 6, cerevislae)-like
Hs.12409 somatostatin
Hs.128087 coagulation factor II (thrombin receptor
Hs.1313 tumor necrosis factor (ligand) superfamily member 8
Hs.133089
Hs.142255 islet amyloid polypeptide
Hs.146355 v-abi Aberson murine leukemia viral oncogene homolog 1
Hs.151988 mitogen-activated protein kinase kinase kinase 5
Hs.159494 Bruton agammaglobulinemia tyrosine kinase
Hs.1613 adenosine A2a receptor
Hs.170917 prostaglandin E receptor 3 (subtype EP3)
Hs.1722 interleukin 1, alpha
Hs.1724 interleukin 2 receptor, alpha
Hs.172894 BH3 interacting domain death agonist
Hs.176090 PRKC, apoptosis, WT1, regulator
Hs.177486 amyloid beta (A4) precursor protein (protease nexin-II, Alzeimer dise...
Hs.180338 tumor necrosis factor receptor superfamily, member 12 (translocating...
Hs.1817 myeloperoxidase
Hs.1846 tumor protein p53 (Li-Fraumeni syndrome)
Hs.2007 tumor necrosis factor (ligand) superfamily, member 8
Hs.20313 protein tyrosine kinase 2 beta
Hs.21486 signal transducer and activator of transcription 1, 91kD
Hs.225995 somatostatin receptor 3
Hs.239176 insulin-like growth factor 1 receptor
Hs.239489 TIA1 cylotoxic granule-associated RNA-binding protein
58 rows (Total: 750 msec)
Abort

FIG. 23B

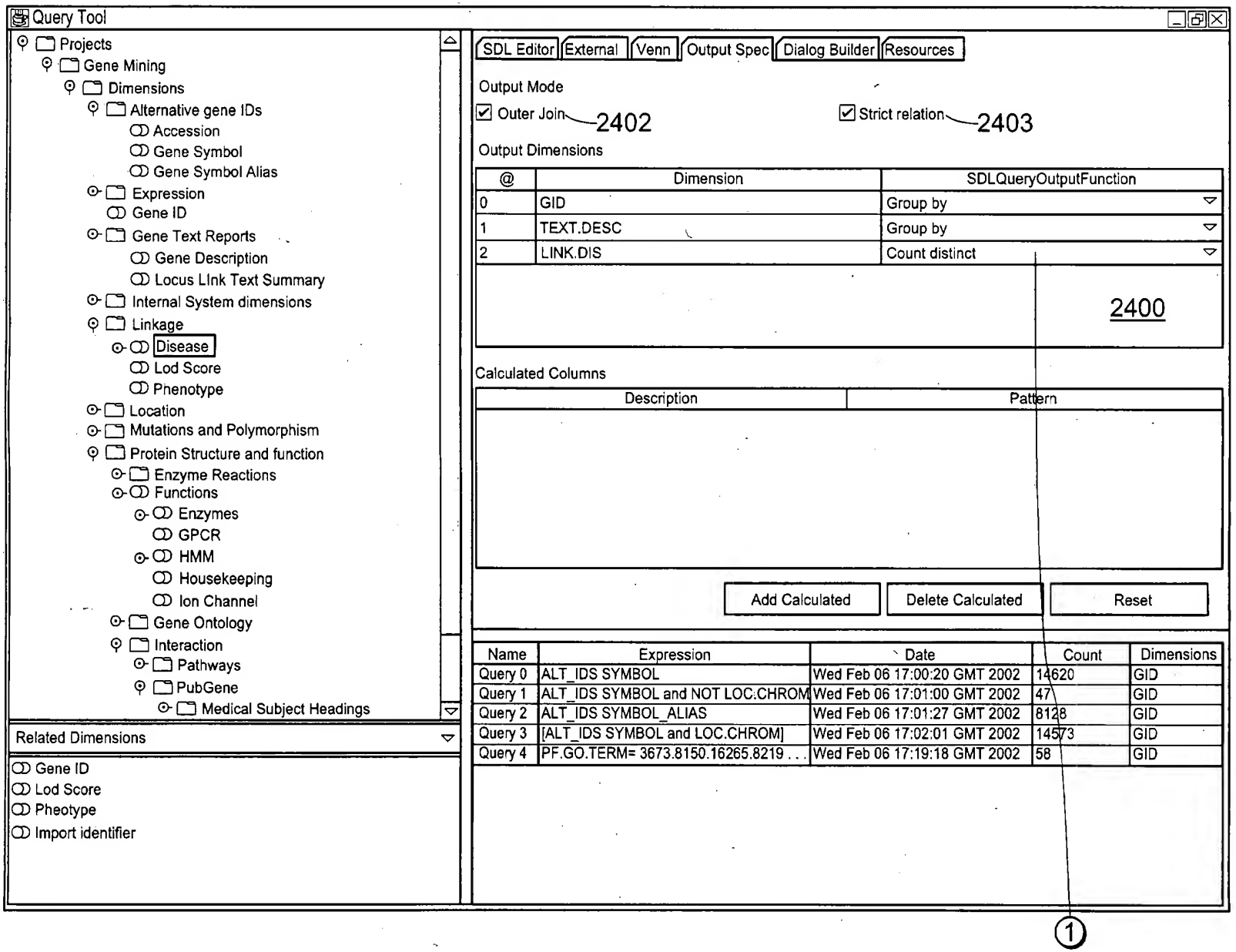

FIG. 24A

FIG. 24B

FIG. 24

FIG. 24A

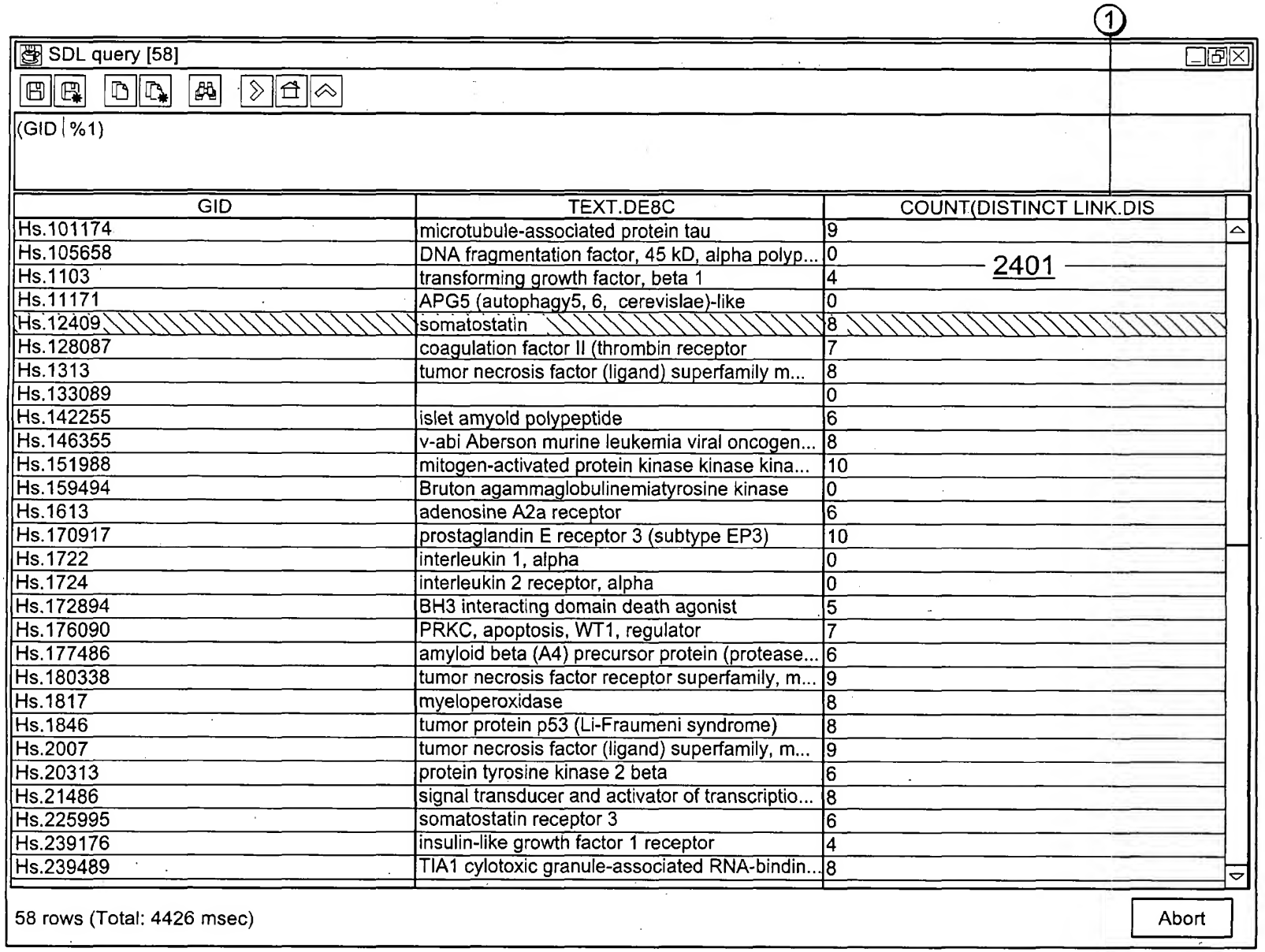
1
SDL query [58]
(GID | %1)
GID
TEXT.DE8C
COUNT(DISTINCT LINK.DIS
Hs.101174 microtubule-associated protein tau 9
Hs.105658 DNA fragmentation factor, 45 kD, alpha polyp... 0
Hs.1103 transforming growth factor, beta 1 4
Hs.11171 APG5 (autophagy5, 6, cerevislae)-like 0
Hs.12409 somatostatin 8
Hs.128087 coagulation factor II (thrombin receptor 7
Hs.1313 tumor necrosis factor (ligand) superfamily m... 8
Hs.133089 0
Hs.142255 islet amyold polypeptide 6
Hs.146355 v-abi Aberson murine leukemia viral oncogen... 8
Hs.151988 mitogen-activated protein kinase kinase kina... 10
Hs.159494 Bruton agammaglobulinemiatyrosine kinase 0
Hs.1613 adenosine A2a receptor 6
Hs.170917 prostaglandin E receptor 3 (subtype EP3) 10
Hs.1722 interleukin 1, alpha 0
Hs.1724 interleukin 2 receptor, alpha 0
Hs.172894 BH3 interacting domain death agonist 5
Hs.176090 PRKC, apoptosis, WT1, regulator 7
Hs.177486 amyloid beta (A4) precursor protein (protease... 6
Hs.180338 tumor necrosis factor receptor superfamily, m... 9
Hs.1817 myeloperoxidase 8
Hs.1846 tumor protein p53 (Li-Fraumeni syndrome) 8
Hs.2007 tumor necrosis factor (ligand) superfamily, m... 9
Hs.20313 protein tyrosine kinase 2 beta 6
Hs.21486 signal transducer and activator of transcriptio... 8
Hs.225995 somatostatin receptor 3 6
Hs.239176 insulin-like growth factor 1 receptor 4
Hs.239489 TIA1 cylotoxic granule-associated RNA-bindin... 8
2401
58 rows (Total: 4426 msec)
Abort

FIG. 24B

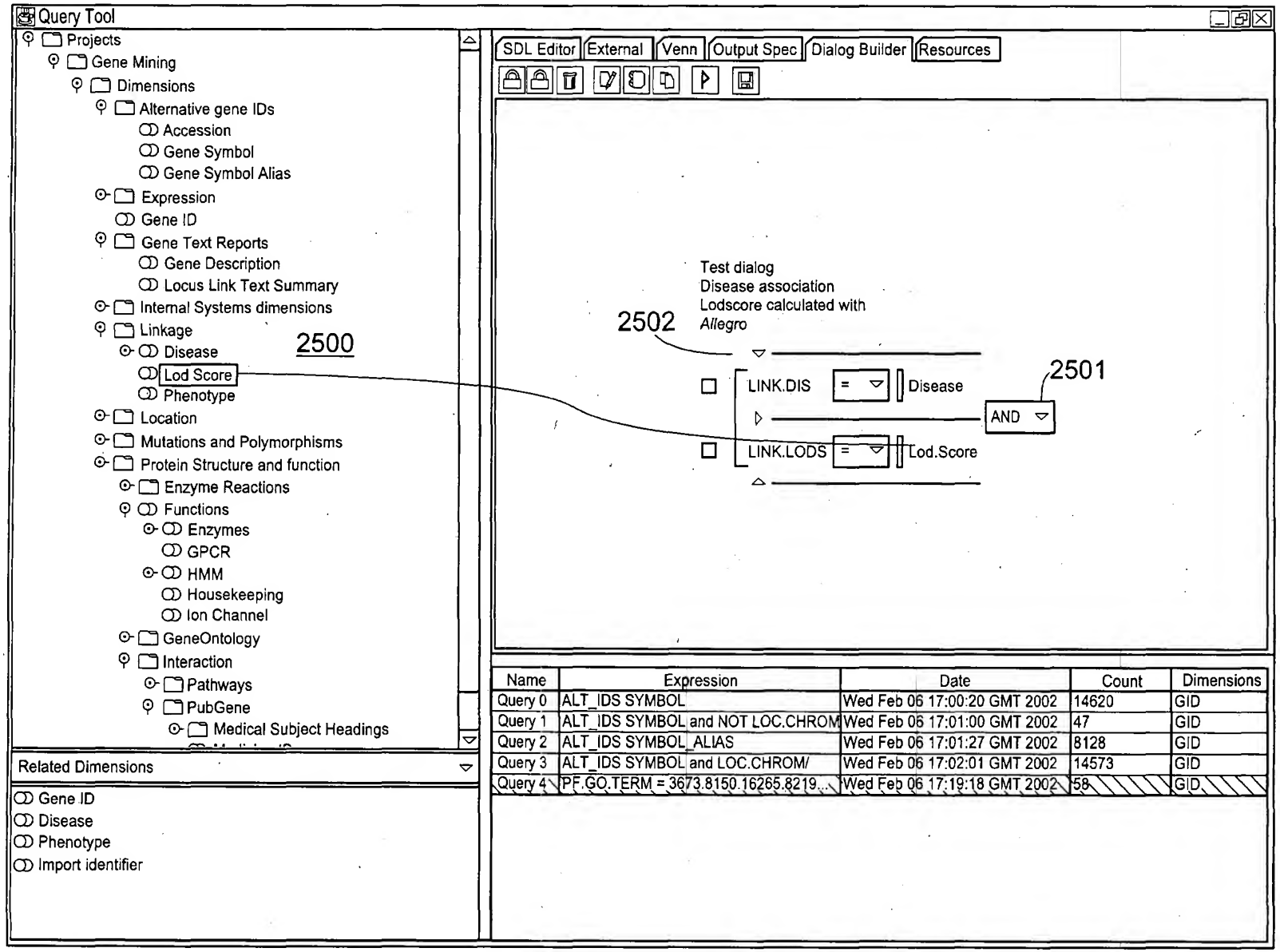

FIG. 25A

FIG. 25B

FIG. 25

FIG. 25A

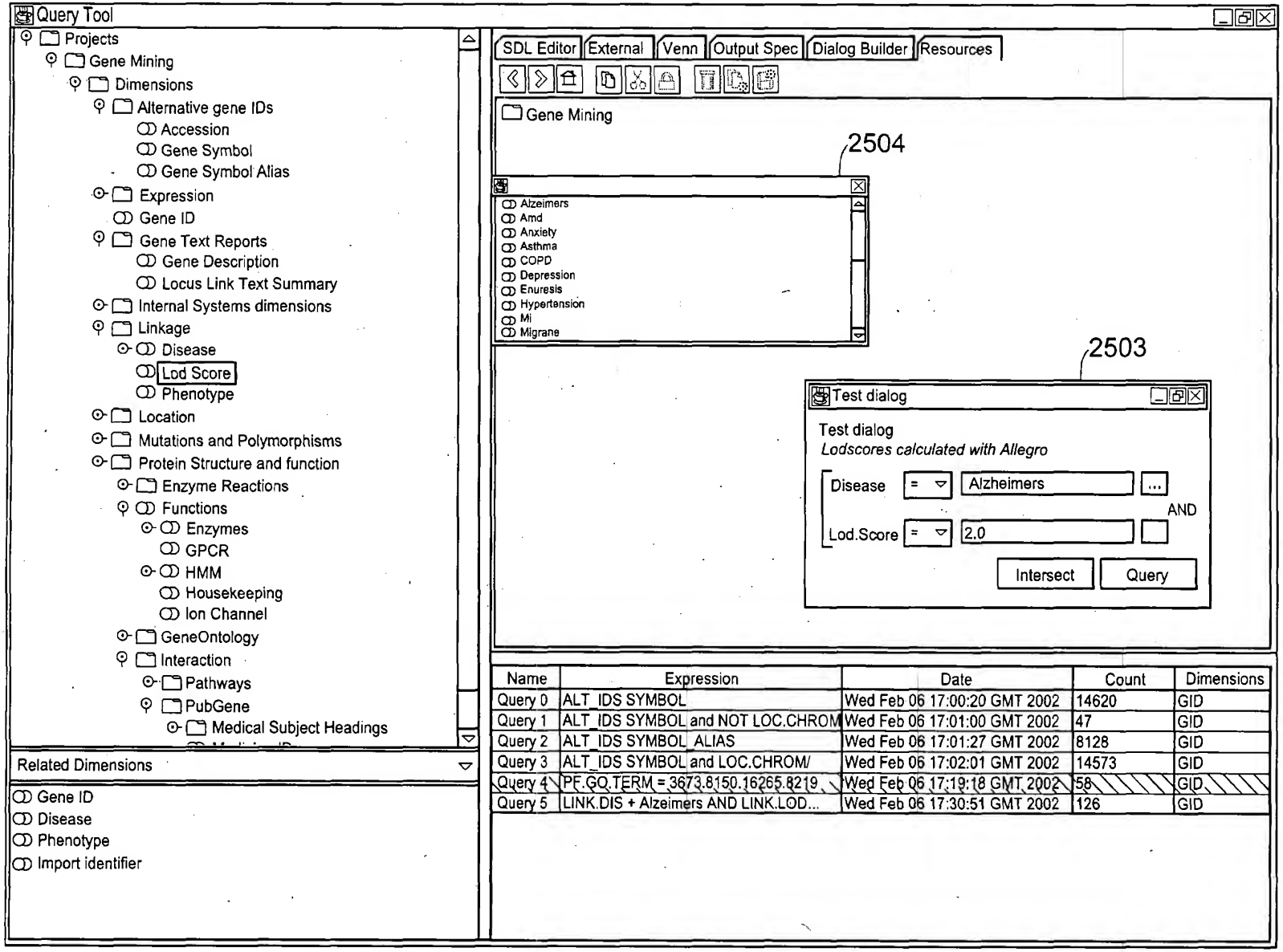
Query Tool
Projects
Gene Mining
Dimensions
Alternative gene IDs
Accession
Gene Symbol
Gene Symbol Alias
Expression
Gene ID
Gene Text Reports
Gene Description
Locus Link Text Summary
Internal Systems dimensions
Linkage
Disease
Lod Score
Phenotype
Location
Mutations and Polymorphisms
Protein Structure and function
Enzyme Reactions
Functions
Enzymes
GPCR
HMM
Housekeeping
Ion Channel
GeneOntology
Interaction
Pathways
PubGene
Medical Subject Headings
Related Dimensions
Gene ID
Disease
Phenotype
Import identifier
SDL Editor
External
Venn
Output Spec
Dialog Builder
Resources
Gene Mining
2504
Alzeimers
Amd
Anxiety
Asthma
COPD
Depression
Enuresis
Hypertension
Mi
Migrane
2503
Test dialog
Test dialog
Lodscores calculated with Allegro
Disease
=
Alzheimers
...
AND
Lod.Score
=
2.0
Intersect
Query
Name
Expression
Date
Count
Dimensions
Query 0
ALT_IDS SYMBOL
Wed Feb 06 17:00:20 GMT 2002
14620
GID
Query 1
ALT_IDS SYMBOL and NOT LOC.CHROM
Wed Feb 06 17:01:00 GMT 2002
47
GID
Query 2
ALT_IDS SYMBOL_ALIAS
Wed Feb 06 17:01:27 GMT 2002
8128
GID
Query 3
ALT_IDS SYMBOL and LOC.CHROM/
Wed Feb 06 17:02:01 GMT 2002
14573
GID
Query 4
PF.GO.TERM = 3673.8150.16265.8219
Wed Feb 06 17:19:18 GMT 2002
58
GID
Query 5
LINK.DIS + Alzeimers AND LINK.LOD...
Wed Feb 06 17:30:51 GMT 2002
126
GID

FIG. 25B

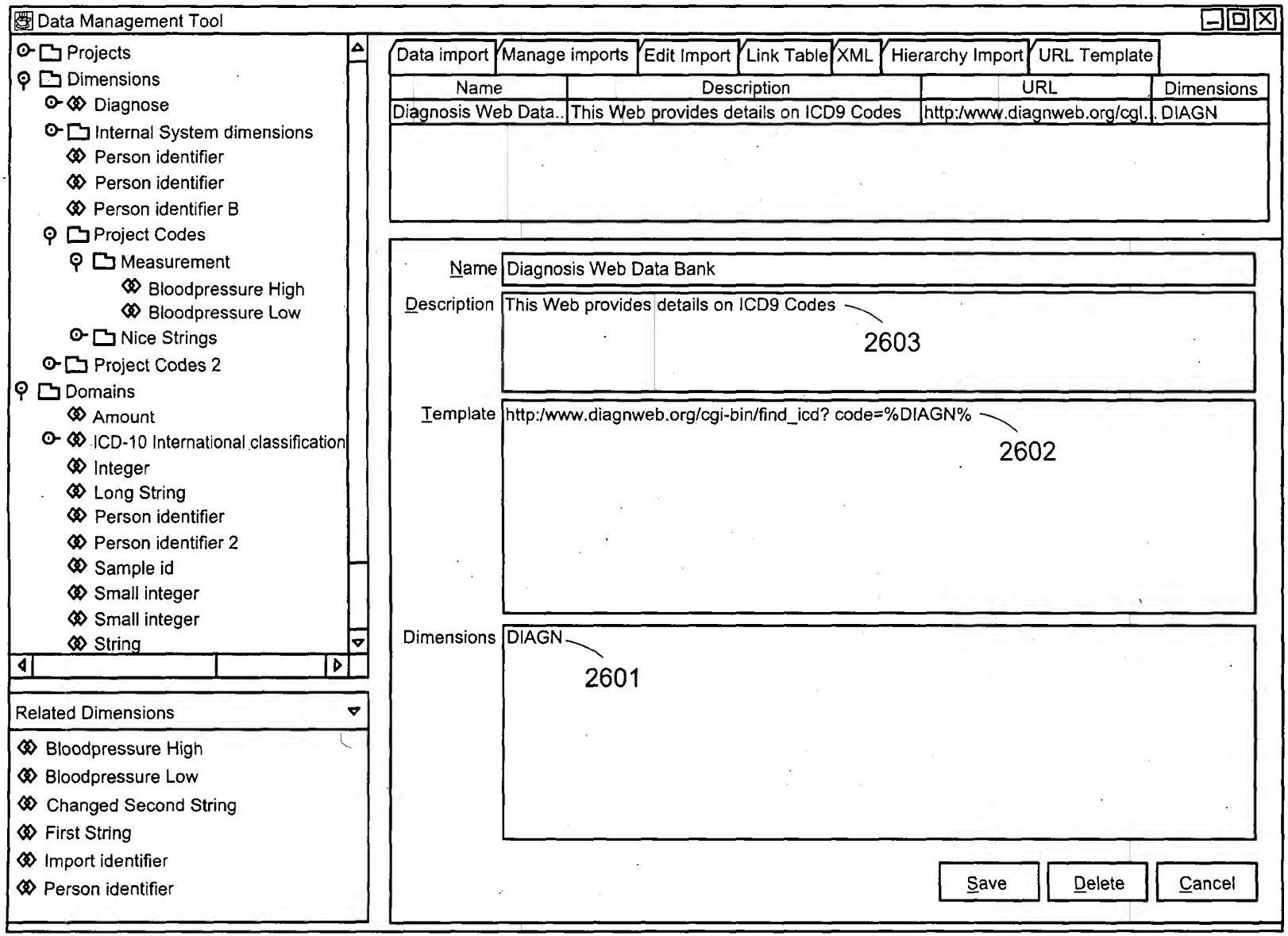

FIG. 26A

FIG. 26B

FIG. 26

FIG. 26A

Docket No. 2345.2041-005

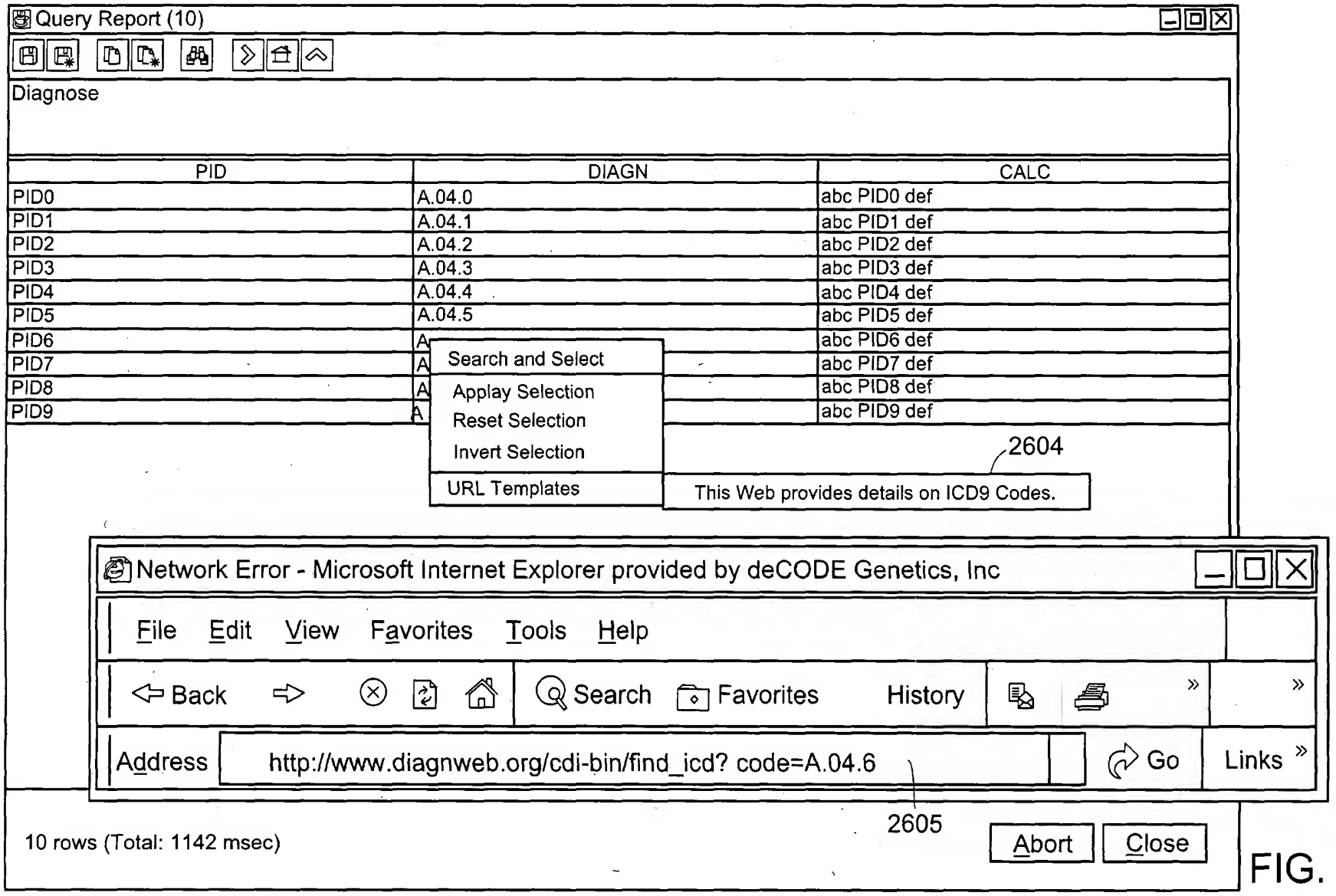
Query Report (10)
Diagnose
PID
DIAGN
CALC
PID0
A.04.0
abc PID0 def
PID1
A.04.1
abc PID1 def
PID2
A.04.2
abc PID2 def
PID3
A.04.3
abc PID3 def
PID4
A.04.4
abc PID4 def
PID5
A.04.5
abc PID5 def
PID6
abc PID6 def
PID7
abc PID7 def
PID8
abc PID8 def
PID9
abc PID9 def
Search and Select
Applay Selection
Reset Selection
Invert Selection
URL Templates
2604
This Web provides details on ICD9 Codes.
Network Error - Microsoft Internet Explorer provided by deCODE Genetics, Inc
File Edit View Favorites Tools Help
Back
Search
Favorites
History
Address
http://www.diagnweb.org/cdi-bin/find_icd? code=A.04.6
Go
Links
2605
10 rows (Total: 1142 msec)
Abort
Close

FIG. 26B

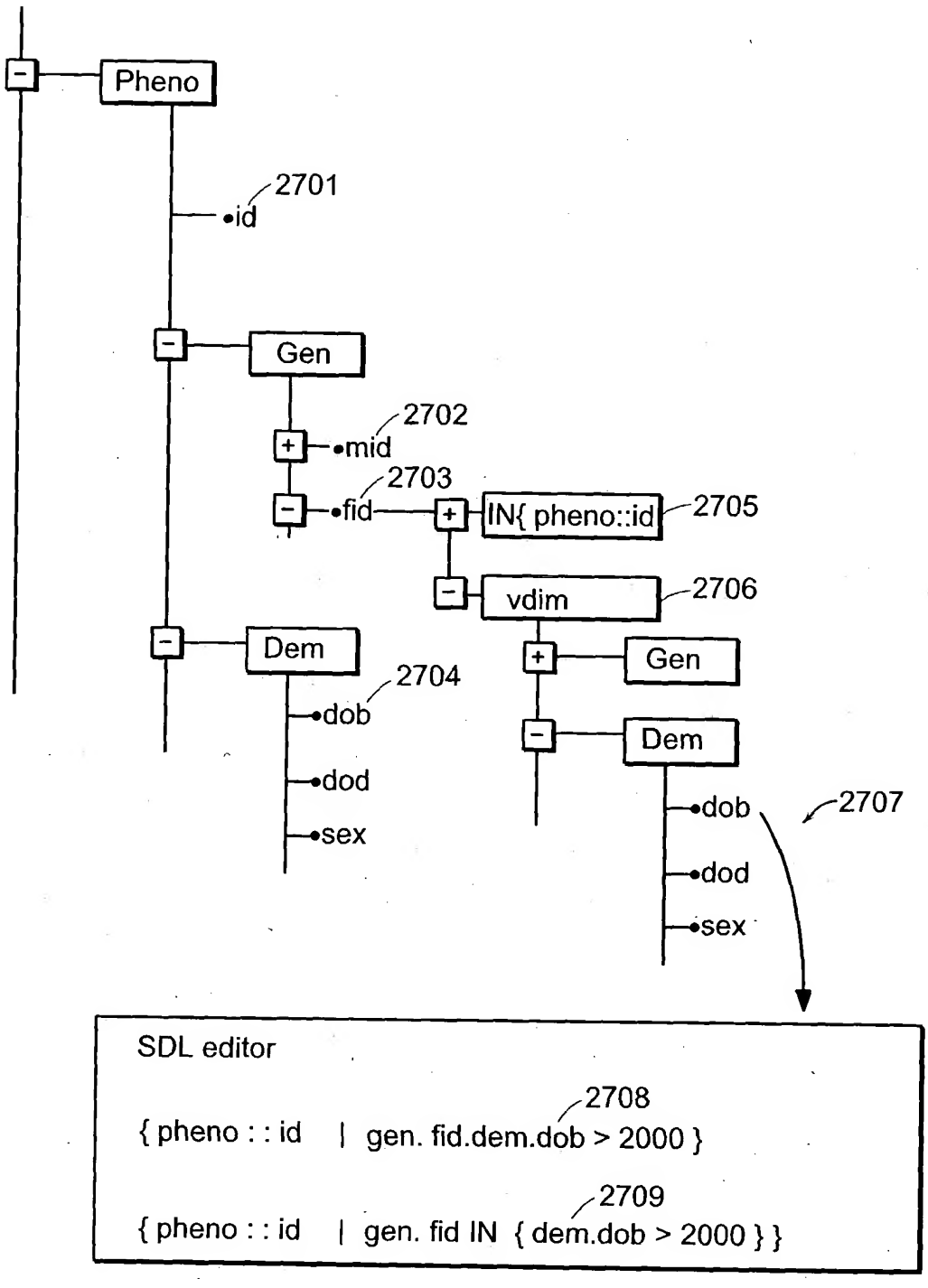
Pheno
2701
id
Gen
2702
mid
2703
fid
IN{ pheno::id
2705
vdim
2706
Dem
2704
dob
dod
sex
Gen
Dem
dob
2707
dod
sex
SDL editor
2708
{ pheno : : id | gen. fid.dem.dob > 2000 }
2709
{ pheno : : id | gen. fid IN { dem.dob > 2000 } }

FIG. 27